Production and Perception of Contrastive and Non-Contrastive Themes in German

Dissertation

zur Erlangung des Grades eines

Doktors der Philosophie

der Philosophischen Fakultäten

der Universität des Saarlandes

vorgelegt von

Bettina Braun

aus Rosenheim

Dekan:	Prof. Dr. Wolfgang Schweickart
Berichterstatter:	Prof. Dr. William J. Barry (Universität des Saarlandes)
	Prof. Dr. D. Robert Ladd (Edinburgh University)

Tag der letzten Prüfungsleistung: 23.07.2004

Zusammenfassung

Die vorliegende Dissertationsschrift beschäftigt sich mit der Frage, wie thematischer Kontrast im Deutschen produziert und wahrgenommen wird. Kontrast im allgemeinen wurde bislang nicht hinreichend definiert; in der Literatur wird Kontrast zum einen oft mit informationsstrukturellem Fokus (d.h. der neuen, noch nicht bekannten Information), zum anderen mit Korrekturen gleichgesetzt. In Kapitel 2 wurden diese verschiedenen Konzepte näher diskutiert und eine Arbeitsdefinition vorgeschlagen: Kontrast ist nur in so genannten kontrastiven Thema–Rhema Paaren eindeutig festgelegt, d.h. in semantisch parallelen Sätzen. Im Kontext einer Frage "Wovon leben die Menschen in Südostasien?" wäre die Antwort "In Indonesien spielt der Tourismus eine große Rolle, und die Malayen leben von der Landwirtschaft" ein solches kontrastives Thema-Rhema Paar. Thematische Information stellt dabei die Verbindung zum Vorkontext dar, während das Rhema neue Information vermittelt. Selbstverständlich ist semantische Parallelität nicht auf Satzebene beschränkt, sondern kann durch einen oder mehrere Sätze unterbrochen werden, wie in folgendem Absatz der Fall. Die semantisch parallelen Sätze sind durch Fettdruck hervorgehoben. Um die Diskursstruktur kohärent zu halten, wurde der Kontrast bereits vor den kontrastiven Satzpaaren durch Gegenüberstellung der Gemeinsamkeiten und Unterschiede angekündigt.

> 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.

Die Realisierung von Sätzen wie z.B. "Die Malayen leben von der Landwirtschaft" in kontrastivem Kontext wurde mit der Realisierung von identischen Sätzen in einem möglichst nicht-kontrastiven Kontext verglichen. Im folgenden Absatz wird "Malaysien" z.B. als Thema eingeführt, welches dann in dem zu analysierenden Satz wieder aufgegriffen wird. In den dazwischenliegenden Sätzen wird zu dem initial eingeführten Thema weitere Information vermittelt. Dieses Abweichen vom Thema und ein Wiederaufgreifen waren notwendig, um eine Pronominalisierung der Vorfeldkonstituente zu vermeiden.

> 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 Internetanbindung. Trotzdem ist Malaysien kein hochtechnologisches Land. **Die Malayen leben von der Landwirtschaft**. Sie sind zwar nicht reich, aber auch nicht besonders arm.

Elf Sprecher, darunter acht Norddeutsche und drei Süddeutsche nahmen an der Datensammlung teil. Sie wurden gebeten, zwölf solche Satzpaare in jeweils kontrastivem und nicht-kontrastivem Kontext laut vorzulesen (diese werden im Folgenden als "Minimalpaare" bezeichnet, ein Begriff, der aus der segmentellen Phonologie stammt). Um die Sprecher davon abzulenken, dass identische Sätze in pragmatisch unterschiedlichen Kontexten vorkommen, wurde eine Vielzahl von Füllmaterial dargeboten. Dies waren ähnliche Absätze, die aber keine "Minimalpaare" enthielten. Darunter waren auch so genannte Ablenkungsabsätze, d.h. Absätze mit einem sehr ähnlichen Thema und mit ähnlicher Diskursstruktur:

> In Malaysia gibt es viel Landwirtschaft. Vor allem Reis wird in diesem tropischen Land viel angebaut. Für den Reisanbau ist ein feuchtes Klima Grundvoraussetzung. Malaysischer Reis ist in ganz Asien beliebt.

Keine der Versuchspersonen bemerkte, dass sie identische Sätze in verschiedenen Absätzen gelesen hatten, was für die Methode spricht. Zwei Leser mussten von der Analyse ausgeschlossen werden, weil sie Probleme mit lautem Vorlesen hatten, sich oft unterbrachen, zögerten und neu ansetzten. Die Äußerungspaare von den übrigen Sprechern wurden akustisch analysiert, wobei der Schwerpunkt auf der satz-initialen Grundfrequenzbewegung lag, dem so genannten thematischen (oder auch pränuklearen) Akzent. Die Grundfrequenz ist das akustische Korrelat der Stimmtonhöhe. Es stellte sich heraus, dass dieser initiale Akzent sehr unterschiedlich realisiert wurde, je nachdem, in welchem Kontext der entsprechende Satz eingebettet war. Insbesondere wiesen die Realisierungen im kontrastiven Kontext folgende Unterschiede in Bezug auf die Realisierungen im nicht-kontrastiven Kontext auf:

- ein höherer Gipfel
- ein späterer Gipfel
- eine größere Dauer des Grundfrequenz-Anstiegs
- ein größerer Umfang im Grundfrequenz-Anstieg
- eine größere segmentelle Dauer der thematischen Konstituente im Vergleich zu den übrigen Elementen in der Äußerung
- ein anderer Rhema-Akzent (bei den Rhema-Akzenten wurde nur die Verteilung der Akzent-*Typen* analysiert). In kontrastivem Kontext gab es signifikant mehr tiefe als hohe Akzente, während diese in nicht-kontrastivem Kontext gleich verteilt waren.

Allerdings wurden nicht alle diese Variablen von allen Sprechern in gleichem Maße benutzt. Es gab Sprecher, die mehr dazu tendierten, den Akzent mit einer höheren End-Grundfrequenz zu produzieren, andere realisierten den Gipfel vornehmlich später und wieder andere machten von beiden Möglichkeiten Gebrauch. Dies lässt den Schluss zu, dass Sprecher verschiedene Möglichkeiten — oder Strategien — verwenden, thematischen Kontrast intonatorisch zu markieren.

Bei den Perzeptionsexperimenten in Kapitel 4 ging es darum, ob Hörer diese akustischen Unterschiede wahrnehmen. Da die reine Diskriminierungsfähigkeit wenig Rückschlüsse auf die Interpretation von thematischen Akzenten zulässt, stand die Frage im Vordergrund, ob Hörer in der Lage sind, die entsprechenden Realisierungen auch dem richtigen Kontext zuzuordnen. In Anbetracht der Sprecherunterschiede lag es nahe, die perzeptuelle Relevanz der einzelnen akustischen Variablen näher zu untersuchen.

Versuchspersonen bewerteten, wie passend eine Äußerung im jeweiligen Kontext klingt (die entsprechenden Kontexte werden unten eingehender beschrieben). Die Bewertung erfolgte zunächst anhand einer 7-Punkt Skala, später mit *magnitude estimation*. Bei letzterer Bewertungsmethode ist die Skala nicht nach oben beschränkt und Versuchspersonen haben die Aufgabe, die Äußerungen relativ zu anderen Stimulus-Äußerungen zu bewerten (Bard et al., 1996).

Um die phonetischen Variablen möglichst optimal kontrollieren zu können, wurde resynthetisiertes Material präsentiert. Als Grundlage zur Resynthese dienten neun Äußerungen, von denen vier als Trainingsäußerungen und fünf als wirkliche Stimuli Verwendung fanden. Syntaktisch waren diese Äußerungen denen im Produktionsexperiment sehr ähnlich. Für jede dieser Äußerungen wurden jeweils vier verschiedene resynthetisierte Äußerungen erzeugt, die sich nur in der Realisierung des thematischen Akzentes unterschieden. In der Manipulation der phonetischen Variablen standen die Höhe und die Position des Akzents (*peak height* und *peak alignment*) sowie die segmentelle Dauer der Vorfeld-Konstituente, d.h. des Thema-Akzentes im Vordergrund. Zur Verdeutlichung dient Abbildung 1, in der die Resynthese für eines der Experimente dargestellt wird.

Die resynthetisierten Äußerungen wurden in einem kontrastiven und einem nicht-kontrastiven Kontext beurteilt[1]. Der kontrastive Kontext entspricht einem semantischen Parallelismus, z.B. "Anna liebt Ferien am Meer, aber Marlene fährt gern in die Berge". Dabei wurde der erste Teil des Satzes schriftlich am Bildschirm gezeigt, gefolgt von dem adversativen Diskurspartikel "aber" und drei Punkten, um die Fortsetzung zu verdeutlichen. Nach Klicken eines Lautsprecher-Symbols am Bildschirm konnten die Versuchspersonen die resynthetisierte Äußerung hören.

[1]Als Kontexte benutzten wir explizit nicht die Kontexte aus dem Produktionsexperiment, um die Einflüsse des Diskurskontextes zu minimieren.

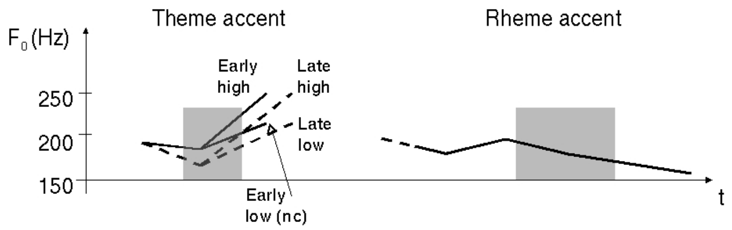


Abbildung 1: Schematische Abbildung von vier resynthetisierten thematischen Akzenten. Die Kontur *Early-low* entspricht der nicht-kontrastiven Realisierung, alle anderen Konturen entsprechen kontrastiven Realisierungen. Der schattierte Bereich kennzeichnet die betonten Silben.

Schriftliche Darbietung:	Anna liebt das Meer, aber ...
Auditive Präsentation:	Marlene fährt gern in die Berge.

Der nicht-kontrastive Kontext wurde durch einen kurzen, visuell dargebotenen Vorkontext und geeignete Kontextfragen festgelegt, die die Vorfeld-Konstituente als möglichst thematisch und nicht-kontrastiv markieren sollten. In folgendem Beispiel steht [$_\theta$ ...] für thematische Elemente, [$_\rho$ ...] für rhematische. Der Satz-Akzent ist mit Großbuchstaben gekennzeichnet.

Schriftlicher Vorkontext:	Ich hab keine Ahnung, was ich Marlene zum Geburtstag schenken soll. Hast du vielleicht ne Idee?
Frage:	Was [$_\rho$ MACHT] [$_\theta$ Marlene] denn gerne?
Antwort (Stimulus):	[$_\theta$ Marlene] [$_\rho$ fährt gern in die BERGE].

Ausgehend von den Ergebnissen des Produktionsexperiments war die Hypothese, dass Hörer in kontrastivem Kontext die stärker markierten thematischen Akzente bevorzugen würden und in nicht-kontrastivem Kontext die schwächer ausgeprägten Akzente.

Im nicht-kontrastiven Kontext hatten die Hörer keine klare Präferenz für bestimmte Versionen, d.h. sie beurteilten alle dargebotenen Stimulus-Äußerungen als gleich passend. Dafür gibt es zwei naheliegende Erklärungen. Zum einen ist es möglich, dass Hörer unbewusst einen Kontrast konstruieren, der ursprünglich nicht in der Situation vorgegeben war. Diese "Reparatur" des Diskurskontexts wird als *accommodation* bezeichnet (cf. Thomason, 1990). Zum anderen ist die Prosodie nicht auf die Markierung von Kontrast beschränkt, sondern kann auch andere Funktionen übernehmen. Im gegebenen Fall könnten stärker markierte Versionen z.B. deshalb akzeptiert werden, weil sie auf eine positivere "Sprecher"-Einstellung schließen lassen als die schwächer markierten Konturen. Diese Hypothesen wurden anhand eines weiteren Experiments überprüft, in dem Versuchspersonen in einem nicht-kontrastiven Kontext zwischen einer stark und einer schwach markierten Version die passendere auswählten. Zusätzlich sollten die Versuchspersonen ihre Wahl zu begründen. Es stellte sich heraus, dass im direkten Vergleich — wie ursprünglich erwartet — die Versionen mit einer schwächer ausgeprägten thematischen Kontur bevorzugt wurden. Die Begründungen dafür waren zum überwiegenden Teil informationsstruktureller Natur (z.B. "das, wonach gefragt wurde, ist betont"). Stark markierte thematische Akzente wurden in einigen Fällen als unpassend bezeichnet, weil sie einen Kontrast herstellten. In den Fällen, in denen die Versuchspersonen unerwarteterweise die Versionen mit einer stärker markierten thematischen Konstituente bevorzugten, wurde dem "Sprecher" meistens eine positive Einstellung attestiert, z.B. freundlich, erklärend, interessiert. Dies legt den Schluss nahe, dass nicht-kontrastive Kontexte intonatorisch unterspezifiziert sind und stark markierten Akzente extra-linguistische Bedeutung beigemessen werden.

In kontrastivem Kontext hatten Hörer eine klare Präferenz für bestimmte resynthetisierte Versionen, woraus zu schließen ist, dass die Intonation hier vorrangig zur Kontrastmarkierung benötigt wird. Versionen mit einem hohen Gipfel wurden am besten bewertet, gefolgt von Versionen mit spätem Gipfel. Dauerunterschiede allein spielten keine Rolle in der Beurteilung. Gipfelhöhe wird im Allgemeinen mit graduellen, paralinguistischen Bedeutungen assoziiert (wie z.B. Emphase) und nicht mit kategoriellen Unterschieden (Ladd and Morton, 1997; Chen, 2003). Daher interessierte mich herauszufinden, ob die Unterschiede zwischen thematischem Material in kontrastivem und nicht-kontrastivem Kontext als graduell oder kategoriell einzustufen sind.

Acht Linguisten, die geübt sind die Satzintonation mit einem bestimmten Annotationssystem zu transkribieren, etikettierten die thematischen Akzente in den beiden Versionen von zehn Äußerungspaaren. Diese zehn Paare wurden zufällig ausgewählt, hatten aber unterschiedliche Gipfelhöhen und -positionen. Beide Versionen eines Satzpaares enthielten aber den gleichen rhematischem Akzent in beiden Versionen. Den Linguisten wurden die zwanzig Versionen isoliert dargeboten und sie sollten den pränuklearen Akzenttyp bestimmen. Bei den meisten Äußerungspaaren wurden die beiden Versionen mit dem gleichen Akzenttyp annotiert ("gleich" bedeutet dabei nicht, dass inter-subjektiv der gleiche Akzenttyp verwendet wurde; vielmehr schien es idiosynkratisch bevorzugte pränukleare Akzenttypen zu geben). Es gab aber auch drei Satz-Paare, deren Versionen von der Mehrheit der Experten mit unterschiedlichen Akzenten annotiert wurden. Die Übereinstimmung in den Akzenttypen war allgemein relativ gering. Die mangelnde Übereinstimmung kann zum Teil dadurch erklärt werden, dass die Unterschiede zwischen der kontrastiven und nicht-kontrastiven Version vor allem im Bereich des Umfangs des Grundfrequenzanstiegs lagen, was mit den GToBI-Akzentkategorien nicht erfasst werden kann. Es könnte also in manchen Fällen in der Tat kategorielle Unterschiede geben, die aber wegen fehlender Akzentkategorien nicht einheitlich annotiert werden konnten. Die beobachtete inkonsistente Annotierung der Daten stellt eine Prosodie–Semantik Schnittstelle in Frage, die eine 1:1 Zuordung von prosodischen Kategorien zu semantischen oder informationsstrukturellen Kategorien annimmt.

Die vorliegende Arbeit beschäftigte sich mit einem Teil-Aspekt der Prosodie–Semantik Schnittstelle, der Realisierung von kontrastiven und nicht-kontrastiven Thema-Akzten. Die Dissertation ist zum großem Teil explorativ und weitere, stärker fokussierte Untersuchungen sind nötig, um die aufgeworfenen Fragen zu beantworten. Diese betreffen nicht nur die prosodische Realisierung von pränuklearen Akzenten in verschiedenen Kontexten, sondern auch die Erforschung der prosodischen Variablen, die zu einem gleichen oder unterschiedlichen Akzent-Perzept beitragen. Erst dann sind vielversprechendere Aussagen zur Prosodie–Semantik-Schnittstelle möglich.

Personal Preface

Many things in life are coincidence. Having left school after my O-levels at the age of 16 and pursued an appreticeship in an electronics company, a PhD degree was not the most obvious thing to come along. It never ceases to surprise me how lucky I was that everything I started seemed to work out. My family is very much down-to-earth and do useful things to earn their living. Even after sitting my A-levels, I was hence a very practically minded person and I would have never imagined myself studying something with no immediate practical value.

So, what brought me to study Phonetics and Phonology? Apart from a general interest for foreign languages, the main impulse came from a charismatic professor who completely changed my way of thinking (for the better or the worse we shall only know in another decade): Bill Barry. He was the one who brought this subject to life, who made it the most interesting discipline in the world, and who was able to inspire others. Apart from teaching the basics in phonetics and phonology, he equipped students with a fair amount of research philosophy and healthy scepticism and made sure we interpreted research results carefully and challenged every methodology. I would like to express my deepest gratitude to him for both his professional and personal guidance, his willingness to chew things over again and again and again, the freedom he allowed me in the choice and direction of research, and his motivation when the end of the world appeared to be coming closer.

Further, I am indebted to Bob Ladd, whom I only met in the last stages of my PhD; he motivated me to bring some Laboratory Phonology aspect into this work. More importantly, he inspired and justified the experiments I was about to begin and enlightened my dark winter months in Edinburgh in 2002/03 with his many ideas and his enthusiasm. He and Bill Barry have always been astonishingly fast in reading and commenting on drafts I sent them. This was very very helpful.

I am also very grateful to Mark Steedman, whom I bombarded with may different drafts on information-structural issues and on the interface between prosody and semantics. I hope these thanks reach him behind the huge piles of papers on his desk.

There have been many others in the last few years who taught me important things, helped me with everyday problems, or motivated me in one way or the other. I can't mention everyone of them, but particularly I would like to thank Jacques Koreman, Jean Carletta, and Frank Keller for helping me to come to grips with statistics, Markus Becker for sharing his gnu-plot experience with me, Christine Haunz for introducing me to magnitude estimation, Verena Helde for agreeing to get my German back into shape for writing the introduction, Stefan Baumann for impersonating GToBI, and Claus Zinn for teaching me how to use cgi-scripts, how to adapt latex-style files and, more importantly, for proof-reading the whole thesis, a great achievement. I am also grateful to Ivana Kruijff-Korbayová and Geert-Jan Kruiff for discussions on the information-structural front. A further big thank you goes to all the students and staff members in linguistics in Saarbrücken (including computational linguistics, phonetics, and psycholinguistics), and everyone involved in the DFG Graduiertenkolleg "language technology and cognitive systems". Their everyday presence, their often critical comments on experimental design and interpretation has boosted this thesis enormously. Needless to say, all the errors in this thesis are my own.

I am very grateful to all those who volunteered to take part in one of the many experiments that were necessary for this thesis; often they gave their time for just a chocolate bar as a "reward". Further I would like to thank Dominik Heckmann, Boris Brandherm, and Lars Peppel for recruiting further participants among their colleagues. Thank you all very much for not letting me down.

Schließlich möchte ich mich bei meinen Freunden und meiner Familie bedanken, die sich viele Klagen anhören mussten und viele meiner Durchhänger mit durchleiden mussten und die mir immer wieder motivierend zur Seite standen.

In any document, there needs to be a serial order, but all your help was equally appreciated and invaluable to me.

Contents

Chapter 1

Introduction

> "Ben stand auf. Ich muß noch was tun. Er sagte das so, als sei ich der, der nichts tun muß." (Timm, 2003, p. 156)
> (English translation: *Ben stood up. "I still have some work to do." He said this in such a way as to imply that I was the one who had nothing to do.*)

In reading this literary quotation aloud, the utterance "Ich muß noch was tun" (*"I still have some work to do"*) is likely to be read first with a neutral intonation (close to the contour shown in the upper part of Figure 1.1). In light of the third utterance (*"He said this in such a way as to imply that I was the one who had nothing to do."*), one will probably pronounce this sentence in a different way because the last utterance indicates that there is a contrast between the two referents (Ben and the literary first person) and between the predications made about them (one has to work, the other doesn't have to). This latter (contrastive) re-reading is probably close to the contour shown in the lower part of Figure 1.1.

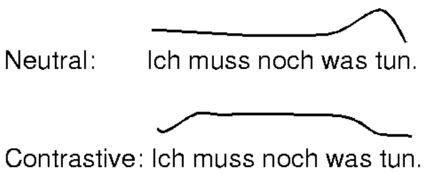


Figure 1.1: Stylisation of a neutral and contrastive realisation of the utterance "Ich muß noch was tun" (*'I still have some work to do'*).

Starting from this literary quotation, we shall motivate what the title of the dissertation "Production and Perception of Contrastive and Non-contrastive Themes in German" comprises. The quotation above indicates that it is possible to perceive this kind of contrast simply by the way it is said. Although, in principle, there are also other means to encode this contrast, such as word order variation or particles, this example shows that contrast can be retrieved from the prosodic realisation only. The introductory example involves two kinds of contrast, one between the respective persons, which are "given" in the situation, and one between the predications that are made about them, which constitute "new" information. In information structural theories, these pre-theoretical notions are termed *themes* (in this situation, the persons) and *rhemes* (the newly provided information).

One main interest of the present thesis lies in investigating the actual acoustic realisation of thematic material produced in contrastive and non-contrastive contexts. Although the intonation contours depicted in Figure 1.1 point to some obvious differences (which are generally associated with the difference between non-contrastive and contrastive themes in the linguistic literature on German), they are stylisations, and actual realisations are likely to vary a lot. We assume that pragmatically distinct contexts (i.e. contrastive and non-contrastive ones) are reflected in the speaker's discourse model, which influences speech planning and hence the actual (prosodic) realisation of utterances. Unlike in the example above we shall analyse utterances with full lexical forms in the thematic constituent instead of pronouns. Pronouns are generally unaccented and accenting them is already very marked, regardless of the actual realisation. Comparing utterances with full lexical forms therefore allows for a more balanced view on intonational realisation.

On the other hand, we aim at exploring how listeners perceive and process differences in the realisation of contrast — provided there are any. One basic assumption is that listeners can judge whether a certain realisation fits a given context (and the corresponding discourse model) or not. This assumption is not trivial, as listeners can extend their context model by adding inferred or presupposed information (a process that is called *accommodation*, see e.g. Thomason, 1990). It is hoped that the insight gained from the production and perception experiments described in this thesis will contribute to a better understanding of the prosody–semantics interface.

This doctoral dissertation is structured as follows:

In **chapter 2** we shall prepare the ground for the experimental chapters to follow. First, we review some annotation schemes for describing intonational form and discuss studies concerned with investigating intonational meaning. We then contrast basic information structure theories, which leads to the discussion of some modern hierarchical approaches. Finally, we propose that contrast is a notion that is independent of information structure primitives. In particular we argue for a narrow definition of contrast which only covers contrastive theme–rheme pairs (semantic parallelism).

In **chapter 3** we summarise the method used for data elicitation and discuss the results regarding the realisations of themes in contrastive and non-contrastive context. We present the acoustic analysis and the statistical evaluation.

In **chapter 4** we discuss a number of perception experiments which are mainly designed as appropriateness ratings. An emphasis lies on exploring the relative importance of the acoustic characteristics of themes in contrastive and non-contrastive contexts.

In **chapter 5** we discuss the implications of the present results for the prosody-semantics interface. With a labelling experiment we investigate whether the data can be annotated consistently and discuss the implications for semantic theories that formalise contrast based on prosodic categories.

Finally, in **chapter 6** we summarise the main experimental results, discuss some theoretical implications, and propose future experiments.

Chapter 2

Background

In this chapter we aim to lay the background for the experimental studies described in chapters 3–5. In section 2.1, our focus lies on presenting annotation systems with respect to the kind of categories they employ to represent the continuously varying fundamental frequency. Furthermore we address how intonational form can be related to intonational meaning. Intonational meaning can be most generally described in terms of information structural concepts. In section 2.2, we therefore discuss basic information structure concepts and more refined, hierarchical theories. In section 2.3 we come closer to the central topic of this dissertation and argue for a notion of *contrast*, which should be distinguished from information structural primitives. Obviously, there is a huge amount of literature on all of these topics, much of which is far removed from the issues which are important here. We have therefore tried to restrict our discussion to material that is of direct relevance for the present thesis.

2.1 Prosodic and Intonational Structure

The adjective "prosodic" comprises a larger range of phenomena than covered with the term "intonational"[1]. Prosodic phenomena include length of segments, intensity and fundamental frequency (f_0), which are perceptually primarily converted into duration, loudness and pitch. In the interpretation of the speech stream, however, these features are used to convey prominence on the word level (i.e. word stress but also speech rhythm) and on the sentence level (i.e. accentuation). The use of the term "intonation", on the other hand, is often

[1] Often, suprasegmental structure is used synonymously for prosodic structure. It should be noted, however, that "suprasegmental" has a very particular meaning coined by the structuralist tradition: Everything that extends the level of single segments (combination of distinctive sounds) is called suprasegmental, e.g. vowel length, elision, assimilation. We therefore use Crystal's definition of prosody, which comprises the "set of mutually defining phonological features which have an essential variable relationship to the words selected" (Crystal, 1969, p. 5).

restricted to the properties of fundamental frequency. For convenience, we shall use the term fundamental frequency (referring to the acoustic phenomenon) and pitch (its perceptual counterpart) interchangeably because confusion seems unlikely.

In this section, we focus on the conversion of the continuously variable intonational (f_0) contour into a symbolic representation on the one hand and on the mapping from intonational phenomena to meaning on the other hand. Any annotation procedure presupposes some kind of interpretation, i.e. an abstraction from the actual physical realisation to assign (meaningful) cognitive categories. There is a lot of evidence for meaningful intonational variation. This is especially obvious in tone languages (e.g. Mandarin Chinese). In tone language, identical lexical items may have different meanings, depending on the tone associated with them. In intonation languages, such as English or German, intonation is not used in a lexically distinctive way. Often, its interpretation is entirely non-grammatical. Intonational form may also directly affect truth-conditional meaning, as shown in example 2.1 (prominence is indicated with small capitals). This utterance becomes false, if Peter introduced someone other than Bill to Mary. In addition, intonation (or more specifically intonational *phrasing*) may signal the internal syntactic structure of utterances, as in the examples shown in 2.2. In the classical reading, there is a semantic parallelism between the propositions about man and God. Using different intonational phrasing (signalling a parenthetic clause), the same lexical string results in a reading in which God thinks that man disposes.

(2.1) Peter only introduced BILL to Mary.

(2.2) Der Mensch denkt, Gott lenkt.
(English translation: *Man proposes, God disposes*)
Der Mensch, denkt Gott, lenkt.
(English translation: *God thinks that man disposes*)

These examples motivate the search for a proper symbolic representation for the description of intonation, so that intonation can interface with syntax and semantics where necessary. On the other hand, intonational annotation should not be too coarse-grained because this could lead to a loss of possibly important information. In section 2.1.1, we first discuss different kinds of annotation schemes, with a special emphasis on autosegmental-metrical (AM) approaches. AM descriptions have been devised for a variety of languages, such as Korean, Japanese, or Greek, and many others are being developed[2]. In section 2.1.2, we then discuss the issue of intonational meaning.

[2]See ToBI-homepage http://www.ling.ohio-state.edu/~tobi/#languages.

2.1.1 Annotation Schemes

Since the mid of the 20th century many different annotation schemes for describing the intonational structure of utterances have been developed[3]. This section is intended to give a rough overview only; more detailed descriptions can be found in various textbooks and articles (Couper-Kuhlen, 1986; Ladd, 1996; Rossi, 2000; Grice and Baumann, 2002, among others). To find one's way through the jungle of different annotation schemes, it is useful to group them according to some criterion. Reviewers of intonation theories often emphasise the distinction between early European, functionally oriented approaches and American, form-based accounts. In the 80s, annotation systems were mainly distinguished with respect to the methods used and the theoretic assumption they relied on. Methodologically, there are annotation systems that are based on an auditory analysis only (e.g. the Dutch IPO-system or the British School), while others use visual information about the waveform (ToBI). As regards the respective theoretical foundations, there are rather phonetic approaches, describing acoustic cues for various functions, such as "finality", "continuation" (Hadding-Koch and Studdert-Kennedy, 1964), or "anger", "surprise", and "boredom" (Williams and Stevens, 1972). These approaches do not assume a mediating phonological structure, but try to map acoustic properties directly to some sort of meaning (for an overview see Lehiste, 1970). On the other hand we find phonological approaches whose preconditions were summarised by Ladd as follows:

> "At a minimum, a complete phonological description includes (a) a level of description in which the sounds of an utterance are characterised in terms of a relatively small number of *categorically distinct entities* — phonemes, features, or the like — and (b) a mapping between such a description and a physical description of the utterance in terms of *continuously varying parameters*" (Ladd, 1996, p. 11)

In our review, a main emphasis lies on the number of "categories" proposed in the various theories and on the phonetic descriptions of these categories. In section 2.1.1 we discuss early level-based approaches (American Structuralism), which describe pitch movement in terms of different pitch levels. In section 2.1.1 we then discuss three contour-based annotation systems (British School, IPO, and KIM), which concentrate on the description of perceptually meaningful pitch movements. Finally, in section 2.1.1, we present modern level-based approaches (a.k.a. autosegmental metrical theories) and discuss some experimental results regarding peak height and peak position.

[3]A summary of earlier intonation descriptions is found in Crystal (1969), chapter 2.2.

American Structuralism: An Early Level-based Approach

The American Structuralist approach to intonation description is worth mentioning for its innovation in using pitch levels to describe continuous pitch movements. This approach to intonation description dates back to Pike (1945) who proposed distinct tonemes for which he postulated four pitch levels. Level 1 represents the highest pitch, while level 4 denotes the lowest. Most later researchers reversed this order (cf. Trager and Smith, 1957); level 4 is now generally used for the highest pitch level. To represent complex tones (such as rises, falls, or fall-rises), combinations between these pitch levels are possible, e.g. 13, 23, 24, 14 to represent rises with different start and end pitch heights and different rise excursions.

This level-based approach, which defines the pitch levels on a relative basis within the speaker's pitch range, was strongly criticised by Bolinger (1951) in his article "Intonation: Levels vs. configurations". There he argued for the perceptual importance of tonal *movements* instead of pitch *levels*. With a level-based system, a locally expanded pitch excursion could easily be captured, see the proposal put forward in the article "Constraints on the gradient variability of pitch range, or, Pitch level 4 lives!" (Ladd, 1994).

Pitch Configurations: British School, IPO, KIM

Judging from the theoretical position Bolinger takes, he can be associated with the British School of intonation[4]. Transcription is mainly performed on a perceptual basis, i.e. without recourse to the actual waveform. In strong opposition to the American structuralist approach, the British School describes tonal *movement* instead of pitch levels and relates the categories identified to particular meanings. O'Connor and Arnold (1961) identified six nuclear tones, a *low fall* and a *high fall*, which only differ in whether the start of the fall is medium or high, a *low rise* and a *high rise*, which again only differ from where the rise starts (low or medium) and two complex tunes, *rise-fall* and *fall-rise*. These six tunes refer to the realisation of the last prominent word (*nucleus*). Everything following the nuclear tone is called *tail* and — per definitionem — cannot contain a prominent word. All tone-group initial syllables up to the first accented syllable are grouped into the *prehead*. The *head* starts with the stressed syllable of the first prominent word. O'Connor and Arnold (1961) identified different combinations of preheads, heads, and nuclei and devised ten different *tone groups*, which are assigned meanings (such as detachment, disclaiming of responsibility). Overall tunes, as opposed to a sequence of tones, have variously been

[4]The British School of intonation dates back to the end of the 19th century, starting with the work of Henry Sweet (Sweet, 1878, 1892). Further researchers in this tradition were Palmer (1922); Kingdon (1958); Armstrong and Ward (1931); O'Connor and Arnold (1961) for the transcription of English, but the approach soon spread to the description of other languages (for German see e.g. von Essen, 1964; Pheby, 1975; Fox, 1984). The British School's approach to intonation transcription emphasises a pedagogical aspect, which originated in the interest in teaching English as a foreign language.

identified as conveying certain meanings, such as "surprise" (Sag and Liberman, 1975) or "uncertainty" (Ward and Hirschberg, 1985). Apart from the meaning associated with holistic tunes, individual tones have also been assigned particular meanings, such as "perfunctory" (Kingdon (1958) for the low rising tone, p. 8), hesitant or apologetic or warning (Kingdon for a low fall-rise, p. 11).

The intonation model developed at the Institute for Perception Research (IPO) was originally intended as a model for specifying intonation for speech synthesis (Cohen and t'Hart, 1967) but has become a general theory of intonation (t'Hart et al., 1990). Sentence intonation is interpreted as a composition of a sequence of pitch movements[5] and connecting line segments. In contrast to the American Structuralist tradition and similar to the British School tradition, only perceptually relevant pitch movements are described. The Dutch system differentiates between prominence-lending pitch movements, which are associated with the lexical stress in a word and non-prominence-lending pitch movements, which occur at the end of intonational boundaries or span a longer stretch of syllables.

Cohen and t'Hart (1967) describe the *hat pattern* as the basic intonation contour. It consists of two prominent pitch movements, in their terms of a "Type 1 Rise" (which occurs early in the accented syllable), and a "Type A fall" (which is again realised early in the accented syllable). They argue that it is the position of the respective movements rather than their slope that plays a role in perception: "the values given for the slopes of rise and fall must not be taken as playing a decisive part in the pattern. It is rather the location of these movements that is highly critical, as could be established from tests for perceptual tolerance by means of the intonator" (Cohen and t'Hart, 1967, p. 187f). The hat pattern may surface both as a pointed hat, in which the two pitch movements are realised on one syllable, or as a flat hat, in which the two movements are separated by a level pitch line segment. IPO also differentiates categories with different pitch-excursions, such as a "Type A Fall" (to the bottom of the range) and a "Type E Fall" (an incomplete fall). The IPO approach is strongly perceptually oriented but in contrast to the British School much research has been devoted to the question of how similar different movements have to be in order to belong still to one category. That is, the IPO tradition uses instrumental methods but they maintain an emphasis on perception.

The Kieler Intonation Model (KIM), which is a less well-know intonation model but of special interest for German intonation, was developed by Kohler and colleagues for the analysis of German (Kohler, 1991a,b, 1997). Like the systems previously described it is

[5]In their 1990 book, t'Hart and colleagues criticise the implicit assumption of modern level-based approaches (as described in section 2.1.1) that "the speaker primarily intends to hit a particular pitch level and that the resulting movements are only the physiologically unavoidable transitions between any two basic levels" (t'Hart et al., 1990, p. 75).

mainly perceptually oriented. The first step in the annotation process is the labelling of four degrees of prominence (level 0 for unaccented, 1 for word stress, 2 for accentuation and 3 for emphasis). Then, with respect to intonational realisation, KIM transcribes terminal patterns, i.e. the intonation contour between the nuclear accent and the intonational phrase boundary. These may be rising (weak or strong rising), falling (terminal, floating, or even) or falling-rising (weak or strong). Finally, three different kinds of pitch accents are annotated, differing only in the position of the peak: early, medial, and late[6]. Kohler (1991b) identified different pragmatic functions for these different accent positions: Early peaks are said to express an established fact, i.e. there is no room for further discussion. Medial peaks signal new facts, which are open for discussion. Late peaks put an emphasis on a new fact and are further described as establishing a contrast to what should already exist in the speaker's discourse model.

In summary, these contour-oriented approaches have a strong perceptual emphasis, which results in the definition of meaningful, perceptually distinct intonation contours. The approaches discussed emphasise different aspects of an intonational contour: The British School describes three different movements, which may have different pitch excursions (*low rise* vs. *high rise*, O'Connor and Arnold, 1961). The IPO approach also has categories to specify differences in pitch range (Type A and Type E fall) but they emphasise the importance of the position of the movement. KIM focuses more on the temporal position of peaks and distinguishes between early, medial and late peaks. Generally, the above approaches maintain a theoretical difference between "nuclear" and "prenuclear" tonal movements (or "tonic" vs. "pretonic" in a different terminology).

Modern Level-based Theories: Autosegmental-Metrical Approaches

Basically, autosegmental-metrical approaches have their roots in the analysis of African tone languages, for which Goldsmith (1976) suggested an autosegmental tonal tier, containing a sequence of high (H) and low (L) tones. The *metrical* aspect was thoroughly explored by Liberman (1975). A further stimulating source was Bruce's analysis of the Swedish word accent (Bruce, 1977), which he analysed as differently localised f_0-targets. That is, words with a lexically defined "word accent 1" have a pitch fall that begins in the syllable preceding the accented one, while the fall begins only in the accented syllable's onset consonant for words with a "word accent 2". The final, integrating step was made by Pierrehumbert's PhD thesis in 1980, in which a finite state model consisting of an initial boundary tone, one or more pitch accents, a phrase accent, and a boundary tone was proposed. This model was

[6]Kohler showed that early and medial peaks are categorically distinct in German.

claimed to be able to account for any possible intonation contour in English[7]. The notion *autosegmental-metrical theory* was only coined in 1996 by Ladd (p. 42) and is standardly used since.

Pierrehumbert (and many subsequent AM analyses) describe the intonation contour as a sequence of (H)igh and (L)ow tones. Tones can be both monotonal (H* and L*) or bitonal (L*+H, L+H*, H*+L, H+L*, H*+H). The diacritic "*" indicates, which of the tones of a bitonal pitch accent is associated with the stressed syllable[8] (examples for the realisation of different pitch accents in German are shown in Figure 2.1). The non- "star-marked" accent is called *leading tone*, if it precedes the starred tone and *trailing tone*, if it follows it. Besides these prominence-lending accents, tones can be associated with boundaries. In Pierrehumbert's system there are minor, intermediate phrase boundaries, signalled by "-" (i.e. H- and L-) and major, intonation phrase boundaries, which cooccur with intermediate phrase accents (H-H%, H-L%, L-H%, and L-L%). A major contribution of Pierrehumbert's model was the specification of phonetic interpolation rules that govern the implementation of the abstract sequence of tones.

In further developments of Pierrehumbert's original model, several changes were made, mainly regarding the tonal inventory (Beckman and Pierrehumbert, 1986; Silverman and Pierrehumbert, 1990), but there were also proposals regarding the status of phrase accents (Grice et al., 2000) and the phonetic reality of interpolation rules (Ladd and Schepman, 2003). The original proposal was significantly simplified, revised, and incorporated into the transcription system ToBI (**To**ne and **B**reak **I**ndices), a labelling standard for English intonation (Silverman et al., 1992; Beckman and Ayers, 1997). The German version of ToBI (GToBI) was originally developed in 1995 and 1996 by Martine Grice, Matthias Reyelt, Ralf Benzmüller, Anton Batliner and Jörg Mayer at the universities of Saarbrücken, Braunschweig and Stuttgart (Reyelt et al., 1996) and has been further improved since. The latest version of GToBI is found in Grice et al. (2005)[9]. Training materials for GToBI are found at: http://www.coli.uni-sb.de/phonetik/projects/Tobi/gtobi.php3.

In Figure 2.1, three types of rising accents in GToBI are shown. They differ with respect to the temporal relation between the stressed syllable (σ^*) and the high and low tonal targets. Grice and Baumann (2002) relate these contours to possible situations and give examples to indicate their pragmatic usage.

Interestingly, modern AM models for the description of intonation languages have rein-

[7]Recently there have been probabilistic models, which show that some combinations of pitch accents and boundary tones are more frequent than others (Dainora, 2002).

[8]Recent experimental studies (e.g. Arvaniti et al., 2000) have shown that the interaction is more complex, see discussion below.

[9]Note that in Stuttgart, a different system for German annotation was devised, see http://www.ims.uni-stuttgart.de/phonetik/joerg/labman/STGTsystem.html.

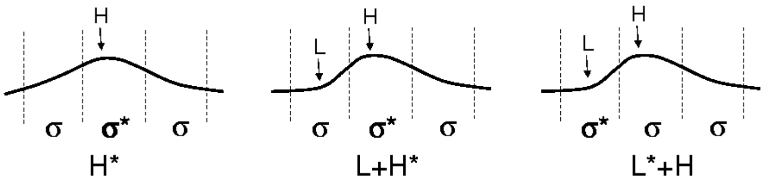


Figure 2.1: Three types of rising accents. The vertical lines represent syllable (σ) boundaries. The stressed syllable is marked by an asterisk.

troduced pitch levels, mainly two (H,L) but sometimes three (H,M,L), see Féry (1993). Following the early work by Bruce (1977), and denying the claims of contour-based annotation schemes like the British School, modern AM theories question the importance of pitch movements: "[...] reaching a certain pitch level at a particular point in time is the important thing, not the movement (rise or fall) itself." (Bruce, 1977, p. 132).

After this general introduction into the autosegmental-metrical framework and the ToBI annotation system we shall discuss further particulars regarding the alignment (temporal position) and scaling (vertical position) of accents. ToBI relies on the assumption that there is a categorical distinction based on differences in peak alignment, which is reflected by different accent types (e.g. L*+H vs. L+H*). As regards pitch height, ToBI and most other autosegmental-metrical approaches distinguish between a high and a low tone, which are categorically distinct. Furthermore, local pitch height differences can be described by means of two diacritics. A phonological narrowing of the pitch range (downstep) is signalled by !H. An expansion of the pitch range is signalled by "upstep" (ˆH). The empirical acoustic investigation and testing of these predictions is tedious because both dimensions are heavily influenced by many different factors, such as speech rate, segmental structure, utterance length, number of accents in the utterance, position of accents and so on.

Scaling of accents. Researchers were initially more interested in investigating the relative nature of pitch height phenomena than they were in exploring the nature of peak alignment (Liberman and Pierrehumbert, 1984; Ladd, 1988). One major problem for any level-based approach is to take into account and to model that a speakers' overall pitch *range* is generally narrowed down towards the end of an utterance (or rather intonation phrase or breath group, the latter already indicating a possible physiological reason for this observation). That is, the notions H and L should not be interpreted as absolute pitch heights, but with respect to the relative position of a particular accent within an utterance (Liberman and Pierrehumbert, 1984; Terken, 1991). Generally, in two subsequent H* accents the latter is scaled lower with respect to the first one, depending on the baseline and the topline. A gradual lowering (and narrowing down) in pitch range is usually referred to as *declination* but the mechanisms and details are still not well understood. Ladd (1996) describes *normalising* models, which interpret accents with respect to an *idealised speaker range* (p. 256). Many different normalising models have been proposed in the literature; for a review see Ladd (1996), p. 256ff. Using the speakers' pitch range as a reference point (or rather *two* reference points, defined by the baseline and the topline) is reminiscent of the intuitive distinctions used in perceptually oriented descriptions, such as in IPO or the British School. Both have tone distinctions, which are only based on the pitch height of the start or the end of the contour (with respect to the speaker's range). In the IPO system, there is for instance a distinction between a "Type A Fall" (to the bottom of the range) and a "Type E Fall" (which is described as an incomplete fall). Similarly, O'Connor and Arnold (1961) distinguish a *low rise* from a *high rise*.

Liberman and Pierrehumbert (1984) investigated the influence of pitch range (the distance between topline and baseline, which can be rather wide or narrow), prominence (triggered by focus-inducing questions) and declination (the narrowing down of the range over time) on the scaling of accents in order to devise a (mathematical) model. They asked subjects to produce a list of berry-names (containing between two to five items), e.g. *'Blueberries, bayberries, raspberries, mulberries, and brambleberries'* in three different degrees of emphasis. They found that in conditions with increased emphasis (i.e. increased pitch register), mainly the fundamental frequency of the H tones was affected but there was only little effect on the scaling of L tones. From their experimental data they conclude that there is a specific "reference level" for the three different pitch range instructions (p. 188).

While Liberman and Pierrehumbert investigated global emphasis, Nolan (1995) investigated scaling effects associated with local emphasis in different positions. Nolan asked subjects to read lists of musical notes, with local emphasis on different positions. He found that peak scaling in the accent following the emphasised one was no different from the neutral list reading. This suggests that local emphasis (or prominence) is indeed a *local* effect, which does not influence the overall downtrend and scaling of accents.

With respect to the perception and interpretation of *individual* accents with different pitch heights, studies point to a relation between pitch height and perceived prominence (Ladd and Morton, 1997).

Alignment of accents. In the introduction to ToBI we presented two accent types that are used to describe a rising accent, L^*+H and L+H^*. These two rises are phonologically distinguished by the temporal alignment of the accentual rise with the segmental structure (alignment). Ideally, an L^*+H differs from an L+H^* only in the temporal alignment of the rising contour, with L^*+H being aligned later than L+H^*, see the stylisations in Figure 2.2.

There are both prosodic and segmental factors that influence the actual alignment of a given accent. Prosodic factors are e.g. the tonal context and the position of the accent in

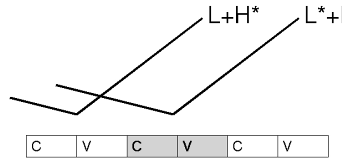


Figure 2.2: Stylisation of the difference in alignment between an L+H* and an L*+H accent. In the simplified segmental structure, “C” is used for onset consonant, and “V” for vowel. The shaded area marks the stressed syllable.

the intonation phrase (cf. Bruce, 1990, p. 107). In the case of “tonal crowding” the peak is aligned earlier because there is another accent close-by (Silverman and Pierrehumbert, 1990; Bruce, 1990). Silverman and Pierrehumbert (1990) also reported that peaks in non-final accents are aligned later in the syllable when the speech rate is higher. On the other hand, the peak is aligned earlier the closer an upcoming boundary is. Segmental effects on alignment were e.g. studied by Rietveld and Gussenhoven (1995). They reported that Dutch falling accents are aligned earlier when the accented syllable ends in a voiceless stop (instead of a nasal).

Pierrehumbert and Steele (1989) investigated whether differences in peak alignment do indeed result in a categorical interpretation or whether these differences are perceived gradually. They synthesised the utterance *‘Only a millionaire’* with a continuous shift of the peak in steps of 20ms. The resulting 15 stimuli were played to subjects, who were told to imitate the melody of the utterances they heard as closely as possible. It turned out that the majority of productions had a relatively early or late peak with only few intermediate realisations. This bimodal distribution of peak positions was interpreted in terms of a categorical contrast between two accent types. Similar results were obtained by Kohler (1991b) using categorical perception tests. He found that in German, early peaks were categorically distinct from medial peaks. He was able to relate the position of the peak to the semantic and pragmatic interpretation of accents (see section 2.1.1).

Recently, there have been many studies on the alignment of accents, which have partly undermined the theoretical assumptions put forward by Pierrehumbert (1980) and Beckman and Pierrehumbert (1986). Studying prenuclear rising accents in Greek, Arvaniti et al. (1998) e.g. found that both the f_0-minimum as well as the f_0-maximum had fixed anchor points but none of them was found within the limits of the stressed syllable. This questions the theoretical foundations of the star-diacritic.

Further AM-Approaches

Recently, alternatives (or rather increments) to the classic ToBI transcription have been proposed. These mainly aim at making the mapping from acoustic features to phonological categories more transparent. These changes are motivated by the low inter-transcriber reliability observed in several studies with ToBI or GToBI transcription (cf. Pitrelli et al., 1994; Grice et al., 1996) on the one hand and the wish to describe more subtle differences than possible with standard ToBI categories on the other hand. In this section we shall briefly present two approaches (IViE and INTSINT), which both introduce an additional phonetic level. This level is to mediate between the physical realisation (f_0 contour) and the symbolic representation (although the mapping to the symbolic representation is not specified for INTSINT yet). It is hoped that such intermediate levels of analysis yield to more adequate and reliable annotation systems in the future.

IViE (Intonational Variation in English). IViE annotation is performed on three tiers (cf. Grabe et al., 1998; Grabe, 2001). First, labellers mark rhythmically strong syllables (P), before the pitch movements in the so-called *intonational domain* (ID) are annotated (the primacy of labelling prominence is reminiscent of KIM). The ID comprises the accent-foot (up to the following accented syllable) and the pre-accentual syllable. IViE offers six labels for phonetic transcription: level pitch or movements on the stressed syllable are transcribed with H, M, or L (levels) or combinations of these basic tones for movements (HL, HM, etc). Further, h, m, and l are used to transcribe the pitch level of the unstressed syllable preceding the stressed one or of unstressed syllables. The phonetic transcription does not supply quantitative information about alignment but should help to limit the range of possible phonological transcriptions of a given intonational event. Motivation for this phonetic representation comes from the extent of regional variation found in English intonation. Grabe assumes that much of the variation is systematic but that phonological transcriptions often loose information about systematic phonetic variation. It is assumed that labeller agreement is higher on this intermediate phonetic level than it is on the phonological level, which allows a consensus to be found or labelling conventions for certain accent types to be specified.

Only after phonetic annotation, classification is performed. The phonological classification is based on ToBI-like accents, but a labeller's choice is restricted to accents for a given variety. The phonological labels differ from standard ToBI accents; particularly, IViE has restricted its tonal inventory to left-headed accents (i.e. there are no accents with leading tones, such as L+H*).

INTSINT (INternational Transcription System for INtonation). INTSINT (Hirst and Cristo, 1998; Hirst et al., 2000) was developed as a means to annotate tonal events on a *surface phonological level*, whose categories can be subsequently mapped onto categories in the *underlying phonological level*. The INTSINT module is part of a larger project with the aim to devise a global theory of intonation. Hirst et al. (2000) argue that a global theory of intonation "will require four levels of analysis: (i) physical (acoustic, physiological) (ii) phonetic (iii) surface phonological and (iv) deep phonological" (p. 51). They emphasise that the representations on specific tiers must be interpretable by adjacent tiers. We focus here on the surface phonological level, on which INTSINT operates. Interpretation of tonal events is both *globally* (i.e. *absolute*, with respect to the speaker's pitch range, such as **T**op, **M**id, **B**ottom) and *locally* (i.e. *relative*, with respect to a preceding target point, such as **H**igher, **L**ower, **S**ame, **U**pstepped and **D**ownstepped). These labels, which are assigned to the turning points of a stylised f_0-contour, are defined on operational acoustic criteria: the highest and lowest target values are e.g. coded with T and B, target points after a pause are coded with M, and all other target points are marked with relative symbols, see example 2.3 (taken from Hirst). INTSINT annotation works automatically, once the intonation boundaries have been manually set.

(2.3)

Il faut que je sois a Grenoble
M T L H L T

So far, INTSINT has no rules mapping surface phonological categories onto underlying ones; so it must be regarded as incomplete. The authors are confident, however, that

> "[t]he automatic generation of large quantities of symbolic data will without doubt make it possible in the next decade to develop a far more complete description of the prosody of different languages than is at present feasible." (Hirst et al., 2000, sec. 4)

IViE and INTSINT, but also recent acoustic work in autosegmental-metrical phonology, show a growing emphasis in the actual fundamental frequency of utterances. It is more and more believed that a pure acoustic analysis of intonational events is fruitful for a better understanding of intonational events and hence a more comprehensive intonational phonology.

2.1.2 Intonational Meaning

For a proper theory of intonational phonology, we do not only need a concise description system for intonational events but also a relation to intonational meaning. In Pierrehumbert and Hirschberg's words, this reads as follows:

> "[...] any theory of transcription must be seen as provisional unless it is supported by considerations both of sound structure and interpretation." (Pierrehumbert and Hirschberg, 1990, p. 282)

Intonation is used for many different purposes, even in intonation languages, which do not use pitch in a lexically distinctive way: Apart from signalling syntactic constituency (as shown in example 2.2), intonation helps to decode different sentence modalities, such as the difference between a question and a statement, e.g. *'Peter is coming (?)'*. Besides, intonation is used to signal non-grammatical functions, such as attitudes, speech acts, emotions (mostly in association with voice quality), emphasis etc. Especially for the non-grammatical uses, a clear categorical distinction seems questionable:

> "I have argued that one of the principal ways in which suprasegmental phenomena differ from the rest of language is that they involve the systematic use of *gradience*. It should be emphasised that this is an aspect of meaning and not merely of form." (Ladd, 1980, p. 203)

Despite the probably gradual nature of intonational meaning, humans are quite good at extracting the extra-linguistic "meaning" and — as Pike pointed out — this additional meaning is quite important to us:

> "the hearer is frequently more interested in the speaker's attitude than in his words — that is, whether a sentence is 'spoken with a smile' or with a sneer [...] If one says something insulting, but smiles in face and voice, the utterance may be a great compliment; but if one says something very complimentary, but with an intonation of contempt, the result is an insult." (Pike, 1945, p. 22)

But strikingly, this important bit of information is not detached from the literal meaning of an utterance but it interacts with it (cf. also Cutler, 1977; Gumperz, 1982; t'Hart et al., 1990).

> "We may regard the words as a rough guide to the meaning, and the intonation as giving greater precision and point, but this is not to say that intonation makes a greater contribution to the whole than does the verbal structure; indeed the intonation without the words would give a very vague impression of the total meaning." (O'Connor and Arnold, 1961, p. 3)

> "Clearly, prosodic cues are systematically based in conventionalised patterns of prosodic usages. However, each prosodic cue we discussed in our analysis did not 'mean' anything in isolation, i.e. by being an instance of a particular type, but by having a signalling value dependent on discourse context and on the previous experience of the listener." (Gumperz, 1982, p. 104)

That is, listeners make use of all available information sources to interpret the meaning of utterances. This mutual influence makes it obviously harder to single out the relative contribution of intonation from the literal meaning of the utterance (excluding intonation). In the remainder of this intonational section we describe the intonational marking of discourse (or more specifically topic) structure and a general theory devising a compositional semantics for AM theories.

Discourse structure. Larger paragraphs are made up of hierarchically organised utterances. The rhetorical relations between utterances can be quite complex, as discourse analyses show (cf. Grosz and Sidner, 1986; Mann and Thompson, 1988). Some of the internal structure of larger discourse is reflected in the intonational structure (cf. Brown and Yule, 1983; Swerts, 1994; Wichman et al., 2000).

Brown and Yule (1983) discussed the notions of *major* and *minor paratones*: major paratones denote the spoken equivalent to the paragraph. Brown and Yule found that the start of these major paratones was signalled by a high pitch onset; their end was marked by a very low pitch together with the presence of a noticeable pause. As regards minor paratones, their initial pitch height could be high, mid, or low to convey different functions. Minor paratones with a high onset are said to initiate major paratones, i.e. introduce a new topic; those with a mid onset are claimed to signal extensions and additions to the previous propositions; low beginnings in minor paratones mark hypotactic subordinations. Their account was mainly perceptual and not accompanied by acoustic analyses.

In his PhD thesis, Swerts (1994) analysed prosodic features that correlated with discourse structure. He found, e.g. that the timing of nuclear falls influences the perceived "finality" of an utterance, with early falls signalling stronger finality than late falls. Similarly, it turned out that the onset of new discourse segments was marked by a longer pause and a pitch reset.

Wichman et al. (2000) analysed the intonational realisation of "initial" elements. Starting from hypotheses derived from extensive corpus studies, they carried out a controlled experiment to investigate the tonal alignment in four different words, which occurred in paragraph-initial position, in sentence-initial position (i.e. the first accent), and sentence-final position. It was assumed that paragraph initial occurrence marks the discourse topic (which is not necessarily the case, see discussion in Wichman, 2000). They found a significant influence of text-position an both peak height and peak alignment: the "more initial" a given accented word, the higher and later its peak.

Compositional semantics. We already discussed that some tones or tunes have been associated with particular meanings: Tones were said to convey e.g. a perfunctory meaning (Kingdon (1958) for the low rising tone, p. 8), or can be interpreted as hesitant or apologetic or warning (Kingdon for a low fall-rise, p. 11). Other researchers assigned meanings to complete tunes, such as "uncertainty" (Ward and Hirschberg, 1985). O'Connor and Arnold (1961) pursued both directions, a semantics for tones (nuclear pitch movement) and a semantic interpretation of tunes (pitch movement of the whole phrase). However, the overall trend was to work towards compositional semantic analyses.

> "A feature of the present description is that is associates a particular communicative value with each of the tones in a way that holds for all occurrences of that tone [...] To achieve this generality of statement, we are forced to gloss the value of the increment [meaning contribution of a tone, BB] in somewhat abstract terms." (Brazil, 1985, p. 104)

Within his theoretical framework, Brazil assigns meanings to the initial key (pitch at the start of an utterance), to the tonic (accent syllable), and to the termination. His "somewhat abstract terms" often refer to information structure notions, such as speaker-hearer convergence (p. 109), i.e. common ground. These notions are discussed in more detail in section 2.2.

Within AM approaches, Pierrehumbert and Hirschberg (1990) were the first to provide a compositional semantics for English pitch and phrase accents and boundary tones. Similar to the proposal by Brazil, Pierrehumbert and Hirschberg's intonational meaning is often related to information structure. They assume that speakers use "tune to specify a particular relationship between the 'propositional content' realised in the intonational phrase over which the tune is employed and the mutual beliefs of participants in the current discourse" (p. 285). They argue that pitch accents convey information about the "status of the individual discourse referents" (p. 286), i.e. different accent types "indicate whether accented items or things predicated of them are to be included among items H [Hearer, BB] believes mutually believed or whether they should be excluded, whether something predicated of these items should be inferable from beliefs H already holds, or whether relationships in which S [Speaker, BB] believes the items participate should be identified by H." (p. 286f). The contribution of phrase accents and boundary tones is less clear, however: Phrase accents are said to "convey the degree of relatedness to preceding and succeeding intermediate phrases" (p. 287), while boundary tones mark the relationships between intonational phrases and indicate discourse segmentation. A general drawback of their account is that intonational meaning is strongly tied to the literal meaning of the respective utterance. In a comment on their paper, Hobbs (1990) simplified the ideas put forward by Pierrehumbert and Hirschberg and provided a uniform but simplistic semantic theory of pitch accents and boundaries.

In Steedman (2000) and Steedman (2002) parts of Pierrehumbert and Hirschberg's description are used and integrated into a theory of information structure that interacts with syntax and prosody. The issue of information structure, which is important for both a thorough understanding of intonational phenomena as well as for defining the notion of contrast, is described in the next section.

2.2 Information Structure

Information structure is a fairly broad discipline. In this dissertation, we can neither cover all approaches presented in the literature so far nor can we supply a complete discussion of the theoretical foundations and coverage of single theories[10].

This section is intended to supply the basic concepts of information structure that are necessary for a proper understanding of the concept *contrast*, which will be motivated and defined in the next section. Information structure, or information packaging, describes how speakers organise information conveyed in utterances. While this organisation of information rarely affects the propositional content of the utterances[11], there is some extra-propositional contribution, which has already been noticed in the nineteenth century (Paul, 1880). In 2.4, both utterance realisations denote that *'Peter gave a book to the boy'* but they highlight different aspects of the message:

(2.4) a) Peter gave a book to the BOY.

b) Peter gave the boy a BOOK.

In 2.4*a* the recipient of the action is highlighted (*'boy'*), while in utterance 2.4*b* it is emphasised that it was a *'book'* that Peter gave to the boy. One way to explain the realisational differences is to infer the contexts in which the examples 2.4*a* and *b* were uttered. Example 2.4*a* would be appropriate in a context where the recipient was questioned, whereas 2.4*b* would be appropriate only if the respective object was under debate. Note that the two utterances are not interchangeable in the respective contexts.

Since utterances are only rarely produced in isolation, they are normally structured in a way to make them felicitous to the context. In cognitively oriented approaches (Chafe, 1976; Prince, 1981; Vallduví, 1990; Vallduví and Engdahl, 1996), it is believed that the speaker takes notice of the mental state of the hearer and therefore tries to relate the current utterance to the information, which he assumes to be already shared by him and the listener. In terms of speech production, this presupposes a rather complex procedure, which Levelt summarised as follows:

> "But the speaker must not only plan and order the contents of successive speech acts. The contents of each speech act should also be brought into perspective, and a particular information structure has to be assigned. The distribution of what should be expressed as topical, focused, or new information must be

[10]Fore more detailed introductions and discussions, see Vallduví and Engdahl (1996); von Heusinger (1999); Kruijff (2001); Kruijff-Korbayová and Steedman (2003). A historical overview is found in Newmeyer (2001).

[11]The propositional content is affected in utterances with so-called focus-sensitive particles, such as *even*, *also*, and *only* (König, 1991). A short discussion is found in section 2.2.2.

assigned, and the speaker will have to acknowledge certain language-specific requirements the message has to satisfy." (Levelt, 1993, p. 109)

In the following section, we try to explain the notions "topical", "focal" and "new information" as theory-independently as possible. The aim to remain "theory-independent" has arisen from the terminological confusion found in the literature:

> "Terms like *contrast, focus, topic*, and *rheme* inhabit a terminological minefield that has hindered research in pragmatics and discourse for decades." (Vallduví and Vilkuna, 1998, p. 80)

2.2.1 Basic Concepts

There is consensus in the literature on information structure that there are two orthogonal dichotomies, which describe different aspects of information structure (cf. Halliday, 1967; Vallduví, 1990; Vallduví and Engdahl, 1996; Steedman, 2000; Kruijff-Korbayová and Steedman, 2003, among others). Modern theories, two of which will be discussed in section 2.2.2, incorporate both dimensions. For a proper understanding of the often confusingly used notions, however, it is worth discussing the basic concepts first, before describing these hybrid approaches. A special emphasis lies on demarcating the respective terminological uses from other misleading uses of the same term.

We introduce the basic terminology using the introductory examples of this section, here repeated as 2.5*a* and *b*. This time, the sentences are presented together with the respective question-context that might have triggered the different realisations. It is generally agreed in the research community that a question context is well-suited to determine the information structure of a given utterance, because the felicity of the resulting discourse can be easily assessed[12].

(2.5) a) Q: To whom did Peter give a book?
A: Peter gave a book [$_\rho$ to the BOY].

b) Q: What did Peter give to the boy?
A: Peter gave the boy [$_\rho$ a BOOK].

Both answers have the nuclear accent (signalled by small capitals)[13] on the constituent that is associated with the wh-word of the question. Note that the pitch accent is realised on one element within the focus-domain only. This element is called the *focus-exponent.*

[12]Note, however, that the question context should be used as a diagnostic test only and not as a means to elicit utterances with different information structures because question-answer pairs naturally favour elliptic answers.

[13]At present, we only indicate the *presence* of a pitch accent, not its *type*.

The informative part of an utterance — signalled by $[_\rho \ldots]$ in example 2.5 — has been named in many different ways:

rheme: This is the notion we shall adopt for the remainder of this thesis, following the use in Firbas (1964); Bolinger (1965); Vallduví and Vilkuna (1998); Steedman (2000). From Steedman (2000), we borrowed the notational convention of squared brackets with the Greek symbol ρ, as used in example 2.5. Compared to the terminology discussed below, the term *rheme* is the one with the least ambiguity.

focus: The term *focus* has been used in the present pragmatic informational sense by Chomsky (1971); Jackendoff (1972); Prince (1981); Selkirk (1984); Rochemont (1986); Lambrecht (1994), among others. Unfortunately, this notion has also often been used to refer to entirely different concepts. For some authors the term *focus* is synonymous to intonational prominence, i.e. *accent* (e.g. Steedman, 2000, footnote 7). Although *focus* and *accent* may coincide in certain cases (see e.g. 2.6*a*), an *accent* on particular syntactic constituents may also give rise to multiple focus interpretations, as shown in 2.6*b*, an example taken from Chomsky (1971). The different focus domains (or rather rheme domains) can be shown by appropriate wh-questions, such as *'What was he warned to look out for?'* or *'What happened?'*. The possibility to construct focus-domains of different sizes is called "focus projection"[14]. Following the proposal by Ladd (1980), 2.6*a* is called "narrow focus", 2.6*b* "broad focus".

(2.6) a) He was warned to look out for an exconvict with a $[_\rho$ RED$]$ shirt.
b) $[_\rho$ He was $[_\rho$ warned $[_\rho$ to look out for $[_\rho$ an $[_\rho$ exconvict $[_\rho$ with $[_\rho$ a red SHIRT.$]]]]]]]$.

In the semantics literature, *focus* is often restricted to the nuclear scope of quantifiers (see Vallduví and Zacharski, 1994, for the distinction between pitch accent, semantic focus, and informational focus). A further use of *focus* is found in the computational linguistics literature for the formalisation of the cognitive status of discourse referents. Grosz and Sidner (1986) e.g. suggest that the referents are ordered on a *focus stack*, with the top-most element being the most salient or activated one (the focus). This use of *focus* is exactly opposite to the informational use of it.

new information: The notion *new information* has been primarily used by Halliday (1967): "the constituent specified as new is that which the speaker marks out for interpretation as non-derivable information, either cumulative to or contrastive with what has preceded" (p. 211). Välimaa-Blum (1988) used a related notion to account

[14] Various rules have been suggested to determine the syntactic constituents that have to carry an accent to allow for such cases of *focus projection*, see Chomsky and Halle (1968); Gussenhoven (1984); Selkirk (1984); von Stechow and Uhmann (1986).

for her Finnish data, namely *NewInfo*. An ambiguous use of *new information* refers to the cognitive status of discourse referents (cf. discussion in Allerton, 1978). "Cognitive status" denotes the familiarity of a referent, whether it has been mentioned before, can be inferred from other concepts (e.g. via a bridging inference) or whether it is brand-new (cf. Chafe, 1974, 1976; Prince, 1981; Lambrecht, 1994). The cognitive status of a referent is generally inferable from the linguistic expressions used. While indefinite noun-phrases denote new referents, definite noun-phrases or pronominalisation are used to refer to *given* referents.

The information structural use of *new information*, however, should be interpreted with respect to the discourse context and is not meant to refer to the absolute newness of referents. That is, even a referent, which has been mentioned several times in the preceding context, may be new information in the information structural use of the term (i.e. new in a given context). This is exemplified with the utterance in 2.7 (cf. Lakoff, 1971):

(2.7) John called Bill a REPUBLICAN // and then HE insulted HIM.

In example 2.7, the referents *'John'* and *'Bill'* are mentioned in the first clause, which is why they can be classified as referentially given in the second clause. This justifies the use of proforms. Despite the givenness of the referents, both proforms are accented, which is rather unusual. If they were unaccented, the two clauses would have to be interpreted as sequential actions, *'John'* being the agent in both of them. The prosodic marking of the pronouns, however, reverses the grammatical relations of the referents: In the first clause *'John'* insulted *'Bill'* and in the second clause the *insulting-relation* is reversed. That is, the sentence only makes sense if *'calling somebody a Republican'* means an insult.

The rhematic information, which was signalled by $[_{\rho} \ldots]$ in example 2.5, supplies new information to the given discourse. There are utterances that only convey new information (all-new utterances), but this is only possible if the context is sufficiently restricted so that their informational contribution can be unambiguously integrated into the preceeding context. That is, an answer such as *'a book'* would be perfectly appropriate in the question context of 2.5*b*. Many utterances supply an anchoring part that informs the listener how to process the newly provided information. Similar to the notational variety found for the rhematic part, there are many different terms to denote the less informative part. The most common ones are listed below:

theme: The notion *theme* is the counterpart to the informative *rheme*, which was discussed above. It has been used by Mathesius (1929); Firbas (1964); Steedman (2000), among others and links the utterance to the preceeding discourse[15].

(back)ground and given information: The less-informative part of an utterance has been termed *ground* by Vallduví and Engdahl (1996) or — closely related — *background* by Chafe (1976)[16]. This notion emphasises the cognitive side of information structuring: "The *ground*, in contrast, is already subsumed by the input information state and acts as an usher for the focus. [...] Sentences have a ground only if the context warrants its use, that is, if the ushering is (thought by the speaker to be) required" (Vallduví and Engdahl, 1996, p. 469).

Halliday (1967) used the cognitive term *given information* — which is "recoverable anaphorically or situationally" (p. 211) — as counterpart to *new information*. In accordance, Välimaa-Blum (1988) employed the term *OldInfo*.

presupposition: The information structural term *presupposition* dates back to Chomsky (1971) and Jackendoff (1972). Prince (1981) as well as Sperber and Wilson (1982) used a similar notion called *open-proposition*. It should be interpreted as "any kind of background assumption against which an action, theory, expression or utterance makes sense or is rational" (Levinson, 1983, p. 168) or in Jackendoff's words:

> "As working definitions, we will use [...] "presupposition of a sentence" to denote the information in the sentence that is assumed by the speaker to be shared by him and the hearer." (Jackendoff, 1972, p. 230)

The information structural use of *presupposition* is sometimes equated with the semantic use of the term, as discussed by Strawson (1950)[17]. As Steedman (personal communication) pointed out, the notion *presupposition* misleadingly equates *ground* and *theme*.

topic: The notion *topic* is associated with the distinction *topic–comment* (Hockett, 1958; Strawson, 1964; Gundel, 1974; Dahl, 1974; Reinhart, 1981) or *topic–focus* (Lambrecht,

[15] Note that the notion *theme* was used in a different way by Halliday (1967) who defines it as sentence-initial: "the theme is assigned initial position in the clause, and all that follows is the rheme" (p. 212). He further assigns a referential character to the theme: "The theme is what is being talked about, the point of departure for the clause as a message" (p. 212). This criterion is normally associated with the notion of *topic* as discussed below.

[16] In more recent work, Chafe discusses information states, which can be active, semi-active, or inactive (Chafe, 1994, p. 53ff).

[17] In the semantics literature, certain expressions are said to trigger presuppositions. A definite noun phrase, e.g., presupposes that its referent actually exists, a factive verb, such as *'regret'* presupposes the truth of its complement phrase. As opposed to assertions, presuppositions are said to *survive* (or remain) in contexts where assertions do not. These contexts include negations, questions, hypothetical assumptions, as well as possibility modals.

1994), which refers to an informationally salient element in the theme. We take some time to discuss this notion because this information split is often found in the literature. We will make clear in section 2.2.2, however, that there is a more general view to marking salient elements in the background. This cognitively inspired division dates back to Paul (1880), who used the notions *psychological subject* and *psychological predicate*. The main difference to the previously stated concepts is that the *topic–comment* split emphasises that messages are *about* something or someone and are uttered in order to increase the listeners' knowledge about the respective referent. A good motivation for singling out *topics* is found in Levelt (1993):

> "If a speaker wants to make some predication about some referent, he had better make sure that the addressee understands which referent is being commented upon. [...] Hence, for the benefit of the addressee, the topic has to be somehow marked in the utterance; it will be encoded as *sentence topic*." (Levelt, 1993, p. 98)

Sentence topics should not be confused with the global notion of discourse topics. The most wide-spread definition is that the topic (i.e. sentence topic) is the part of an utterance, the sentence is *about* (Kuno, 1972; Gundel, 1974; Reinhart, 1981; Lambrecht, 1994). Lambrecht emphasises that *subjects* and *topics* are distinct notions. Strawson (1964) defined a "Principle of Relevance", according to which a topic needs to be "a matter of standing current interest and concern" (p. 79).

Topics can be identified with appropriate context questions, such as *'What about x?'* to establish x as sentence topic. Further tests are the *as-for* test (2.8*a*) and the *aboutness* test (2.8*b*). Both examples are taken from Lambrecht (1994), p. 149:

(2.8) a) As for Rosa, John didn't really love her.
b) John said about Rosa that he didn't really love her.

This view presupposes that *topics* are referential; otherwise it would make little sense to make a predication (*comment*) about them. Sentences may be *topic-less*, such as the utterance *'There is a fly in my tea'* (Reinhart, 1981). Sentences of this kind have to be interpreted as *all-focus* (Vallduví and Engdahl, 1996) or *all-new* (Schmerling, 1976); they are uttered *out of the blue* or in the context of an unspecified question, such as *'What happened?'* or *'What's new?'* The issue of referentiality becomes more difficult when temporal or spacial adverbials (such as *'in January'* or *'in Paris'*) are concerned. At first glance they appear to be non-referential. There is cross-linguistic evidence, however, that these can function as topics: Japanese, which has a designated topic-marker ("wa"), allows for topical adjuncts[18].

(2.9)	*Kinou*	*wa*	*eiga*	*o*	*mita.*
	yesterday	TOPIC	movie	ACCUSATIVE	saw

'Yesterday I saw a movie.'

The appropriateness of example 2.9 can be explained if we use an expanded (or wide) definition of referentiality, which includes temporal and spatial adverbs[19]. Further evidence is supplied by the possibility to have linguistically reduced forms for these adverbials (e.g. *'I go to Paris'* vs. *'I go there'*), which indicates that temporal and spatial information can be both new and old (in the sense of Chafe, 1974, 1976; Prince, 1981).

Halliday (1967) defines topics (which he terms *theme*) on a purely structural basis and claims that they have to be sentence-initial. This claim has been variously made also by other authors (Paul, 1880; Firbas, 1964; Allerton, 1978; Reinhart, 1981), and is broadly true for languages with very free word orders, such as Russian or Czech. Such a structural constraint is, however, not accepted by all authors (Lambrecht, 1994; Vallduví and Engdahl, 1996), especially, because in languages such as German or English, focal rhematic material can be positioned sentence-initially ('PETER *made that remark'*). Another issue of debate is the question about the cognitive status of topics, whether they need to be given referents or whether they can be new (terminology according to Chafe, 1976; Prince, 1981). According to Halliday (1967) and Reinhart (1981), topics need not be given, i.e. the predication about a topic can be processed by the hearer even if he has never heard of the referent before. Lambrecht (1994), on the other hand, argues that comments about unidentifiable referents are unacceptable.

Summary

On the whole, the notion of topic seems rather vaguely defined. Nevertheless does it acknowledge a prominent element in the thematic, backgrounded part of an utterance. Different information splits emphasise different aspects of an utterance. It has long been argued that there is valuable information in both distinctions, theme–rheme and topic–comment (Dahl, 1974; Välimaa-Blum, 1988; Vallduví, 1990; Vallduví and Engdahl, 1996; Steedman, 2000). In this section we shall briefly discuss the information structural divisions predicted by theme–rheme and topic–comment partitioning by means of two examples, 2.10 and 2.11.

[18] Note that Japanese is a subject-drop language, i.e. the subject can be unexpressed if the context supplies the relevant information.

[19] Such an expanded definition is in accordance with the formalisation presented in Kamp and Reyle (1993). They treat the utterance time and temporal adverbs in a similar way as discourse referents in their discourse representation theory (DRT).

These examples are adapted from Vallduví and Engdahl (1996). The first context-setting question (*'What about John?'*) highlights the topical constituent. To force different information packagings, a further focus-inducing question is used that highlights particular elements in both the thematic or the rhematic part. Highlighted elements in the thematic part are signalled by italics, highlighted elements in the rhematic part by small capitals. Answers *a* show the predicted theme–rheme structure ($[_\theta \ldots]$ signals the theme, $[_\rho \ldots]$ signals the rheme) that does not assign a special status to referential topic elements. Answers *b* show the topic–comment structure ($[_T \ldots]$ signals the topic, $[_C \ldots]$ delimits the comment).

(2.10) What about John? What does HE drink?

a) $[_\theta$ *John* drinks] $[_\rho$ cooled BEER]

b) $[_T$ *John*]$[_C$ drinks cooled BEER].

(2.11) What about John? What does he DRINK?

a) $[_\theta$ John *drinks*] $[_\rho$ cooled BEER]

b) $[_T$ John]$[_C$ *drinks* cooled BEER].

Independent from the actual realisation of the second focus-inducing question, the theme–rheme structure delimits the newly provided information (*'cooled beer'*) from the anchoring information (*'John drinks'*), while topic–comment splits the referent *John* from the predication of the utterance (*'drinks cooled beer'*). Note that all sample answers in 2.10 and 2.11 have a primary intonational prominence on *'beer'*. Furthermore, a secondary intonational prominence is found on either *'John'* or *'drinks'*, depending on the focus-inducing question. While the topic–comment structure is well-suited to single out the topical element in 2.10*b*, it is unable to account for the secondary prominence (*'drinks'*) in example 2.11*b*. The theme–rheme division is hence more general as it allows for emphasised elements in the theme — no matter whether they are referential or not. A further disadvantage of the topic–comment partitioning is that the newsworthy information does not stand out from the overall predication: In contrast to the rhematic part, which identifies the newly supplied information only, the comment comprises the whole predication about a given referent. In the reminder of this thesis we shall therefore focus on the more powerful theme–rheme divisions.

2.2.2 Modern Hierarchical Theories

In this section we shall introduce two modern hybrid theories (Vallduví and Engdahl, 1996; Steedman, 2000). Both of them postulate two information structure layers that differ in their theoretical foundations. Most importantly, the approach of Vallduví and Engdahl

(1996) partitions on the level of *form* whereas Steedman (2000) partitions on the level of *meaning*. Furthermore, the approach of Vallduví and Engdahl (1996) is cognitively inspired and can be related well to discourse theories, such as *file change semantics* (Heim, 1982) or *discourse representation theory* (Kamp, 1981; Kamp and Reyle, 1993). The approach of Steedman (2000), on the other hand, aims at integrating syntax and semantics using combinatory categorical grammar[20].

The hierarchies used in the two approaches are contrasted in Figure 2.3. Note that both theories share the primary dichotomy — *ground–focus* vs. *theme–rheme* — which is merely named in a different way. A further commonality is that both theories allow for discontinuous information structural primitives (see example 2.14). The theories differ, however, in (the definition of) their secondary partition. The concepts denoted by the respective information structural primitives are discussed below.

ground		focus
link	tail	focus

theme		rheme	
focus	background	focus	background

Figure 2.3: Information structure partitions according to Vallduví and Engdahl (left-hand side) and Steedman (right-hand side).

Vallduví and Engdahl (1996)

Vallduví and Engdahl (1996) partitioned only the background further into a topic-like element called *link* and a genuinely uninformative part called *tail*. They describe their approach as *instruction-based* (p. 468) and refer to a model of information states, which are updated by subsequent utterances. They use a file-card metaphor to illustrate the notions they use: While the *focus* represents the actual update potential, *links* indicate the address of the file-card, i.e. where the new information (*focus*) is to be saved. *Tails* indicate how the information fits into the actual information state. That is, within the backgrounded, thematic material, there are referential (*link*) and non-referential (*tail*) elements. To illustrate their approach, we repeat an example without *tail*, which is discussed in Vallduví and Engdahl (1996), p. 470 and 475 (here shown as 2.12). The part in italics represents the *link*, which is comparable to the *topic*[21], the rhematic part (*focus* in their terms) is signalled by [$_F$ …].

(2.12) Tell me about the people in the White House. Anything I should know?
The president [$_F$ hates CHOCOLATE].

[20] The Prague School similarly advocates hierarchical information structure partitions (Hajičová et al., 1998). Because their concepts are not primitive notions but derived from contextual boundness and communicative dynamism we will not discuss this approach further here.

[21] Note that Vallduví and Engdahl (1996) allow *links* to appear in non-initial positions as well.

We shall discuss the representation of example 2.12 in terms of DRT (see 2.13)[22]. In the left DRS of 2.13, x and y in the upper part denote the discourse referents. Their properties (being president and being chocolate) and the relation between referents are listed in the lower part of that DRS. If we now assume a follow-up utterance such as *‘He is [$_F$ allergic] to it’* with a focus-tail structure, the pronouns are bound to appropriate referents ($a = x$, $b = y$). The resulting DRS is the right one in 2.13:

$$(2.13) \quad \begin{array}{|l|}\hline x,y \\ \hline president(x) \\ chocolate(y) \\ hate(x,y) \\ \\ \hline \end{array} \qquad \begin{array}{|l|}\hline x,y,a,b \\ \hline president(x) \\ chocolate(y) \\ hate(x,y) \\ a{=}x, b{=}y \\ allergic(a,b) \\ \\ \hline \end{array}$$

Sentences may be structured without a *link* if the speaker assumes that the hearer knows where to store the new information. A *link* is essential, however, in the case of a topic change or a contrastive topic (Vallduví and Engdahl, 1996, p. 473).

The analysis of Vallduví and Engdahl (1996) was inspired by Catalan, in which the accent has to be realised on the final word of the core clause. To encode different information structures, Catalan uses syntactic detachments to remove non-focal material from the default accent-position. Interestingly, *link*-material is moved to the left, whereas *tail*-material is detached to the right. In their article, Vallduví and Engdahl (1996) describe the realisation of the information structure primitives used for English, Dutch, German, and Turkish. For English, Vallduví and Engldahl follow Jackendoff (1972) and argue that the focal part is realised with an A accent (which is comparable to autosegmental H*) and the link with a B accent (roughly comparable to autosegmental L+H*)[23]. For German, Vallduví and Engdahl (1996) refer to Féry (1993) and propose a bitonal falling accent (H*+L) as focus accent. Links may be either signalled by a rising tone (L*+H) or a falling one (H*+L).

[22] Vallduví and Engdahl (1996) explicitly relate their information structural approach to a file-card metaphor, which can be formalised in terms of the *file-change semantics* proposed by Heim (1982). We prefer DRT here over *file change semantics* because DRT allows us to represent the whole discourse in a single *discourse representation structure (DRS)*. In file-change semantics, the discourse information is spread over many file-cards.

[23] It seems as a simplification to reduce the accent types to a rising accent (B-accent or L+H*) and a falling one (A-Accent or H*). In autosegmental-metrical approaches, six to seven different types of pitch accents have been proposed (Pierrehumbert, 1980; Beckman and Ayers, 1997; Grice and Baumann, 2002), see section 2.1. It seems hardly possible that only the above mentioned accent type distinctions should be used to signal information structure.

Steedman (2000)

Steedman, in contrast, relates information structure directly to logical form, which is discussed in more detail below. For illustration we repeat one of his examples in 2.14 (adapted from Steedman (2000), p. 459). Round brackets signal the primary dichotomy between *theme* and *rheme*. According to Steedman, the accented items within both *theme* and *rheme* are called *focus* as opposed to *background*. In more recent work, Steedman (2002) proposed the notion *kontrast*, following a proposal by Vallduví and Vilkuna (1998). He claims that thematic kontrast is realised by an L+H* LH%, rhematic kontrast by an H* LL% (note that kontrast is used synonymously to accent by Steedman)[24].

(2.14) Q: I know that Marcel likes the man who wrote the musical.
But who does he admire?

$$\text{A: } (\underbrace{\underbrace{\text{Marcel}}_{\textit{background}}\ \underbrace{\textsc{admires}}_{\textit{kontrast}}}_{\textit{theme}})\ (\underbrace{\underbrace{\text{the woman who}}_{\textit{background}}\ \underbrace{\textsc{directed}}_{\textit{kontrast}}\ \underbrace{\text{the musical}}_{\textit{background}}}_{\textit{rheme}})$$

Steedman (2000) partitions both theme and rheme further; with his information split, prominent parts in both theme and rheme can be distinguished from non-prominent parts. The prominent parts are related to a contrast in the pre-context, as shown in example 2.14.

Formalisation. Before we discuss how Steedman formalises kontrast in both theme and rheme, we briefly motivate why information structure has become important for semantic theories at all. While accentual differences normally do not result in truth-conditional changes (see 2.15*a* and *b*), accentual differences in utterances with so-called focus-sensitive particles do (2.15*c* and *d*):

(2.15)
a) Peter introduced BILL to Sue.
b) Peter introduced Bill to SUE.
c) Peter only introduced BILL to Sue.
d) Peter only introduced Bill to SUE.

In terms of model-theoretic semantics, the utterance in 2.15*c* is true only if *'Peter'* introduced *'Bill'* to *'Sue'* and if he introduced nobody else than *'Bill'* to *'Sue'*. In all other cases the sentence is false. This has been called exhaustive identification (cf. Kiss, 1998b), which is discussed further in section 2.3.1. We will not discuss the formalisation of examples

[24]Steedman (2002) has recently integrated a greater variety of pitch accents, depending on whether the speaker believes the respective information to be mutually agreed or not. Generally, non-agreed information is realised with a low tone: For theme accents, L+H* is the agreed theme, whereas L*+H is non-agreed. For rhemes, H* is the agreed rheme, whereas L* is non-agreed.

such as 2.15*c* and *d*, which often state rules for the semantic composition between *only* and the focal constituents (called *association with focus*, cf. Rooth, 1992).

Following the initial approach by Jackendoff (1972), several formalisations have been developed, such as the *structured meanings* account (von Stechow, 1990; Krifka, 1993) and *alternative semantics* (Karttunen and Peters, 1979; Rooth, 1985, 1992). Despite the differences in detail, whose description and discussion is beyond the scope of this thesis, there is a large common ground. Good semantic overviews on focus semantics are found in Kratzer (1991); von Stechow (1991) and Kadmon (2001).

For discussing the formalisation of Steedman's approach we use example 2.16. Following Steedman, L+H* LH% signals the kontrast within the thematic part, whereas H* LL% signals the kontrast within the rhematic part:

(2.16) Q: I know which result Marcel PREDICTED. But what did Marcel PROVE?

A: (Marcel	PROVED*)*	*(*COMPLETENESS*)*
	L+H* LH%	H* LL%

Following Jackendoff (1972), Steedman formalises the theme (which Jackendoff called *presupposition*) via functional abstraction, i.e. the accented constituent is removed from the semantic formula and replaced by a logic variable x. This is indicated by the symbol "λ" in formula 2.17, which is said to bind the variable x:

$$\lambda x[proved'(x, Marcel')] \tag{2.17}$$

The underspecified formula in 2.17 can be reconverted into a closed formula by combining it with an element for x, such as *'completeness'*. This would result in the proposition $proved'(Marcel', completeness')$. Steedman performs an existential closure on 2.17. This is a very strong semantic claim because it demands that there actually *exists* (∃x) some x that Marcel has proven.

$$\exists x[proved'(x, Marcel')] \tag{2.18}$$

The presuppositional skeleton in 2.18 can be expanded to a set, with the variable x instantiated dependent on the context (Steedman's *rheme alternative set*). This is related to the notion of "alternative sets" introduced by Rooth (1985).

$$\text{rheme alternative set: } \left\{ \begin{array}{l} \text{proved'(decidability'Marcel')} \\ \text{proved'(soundness'Marcel')} \\ \text{proved'(completeness'Marcel')} \end{array} \right\} \tag{2.19}$$

For the two intonation contours, Steedman identifies the following functions:

theme tunes *presuppose* a rheme alternative set
rheme tunes *restrict* a rheme alternative set

The theme tune does not only presuppose a rheme alternative set but also a *theme* alternative set (which is related to Q, the question alternative set in Rooth, 1985; Büring, 1997). The theme alternative set evokes alternatives for the accented element of the theme. A possible theme alternative set for example 2.16 is shown in 2.20:

$$\text{theme alternative set: } \left\{ \begin{array}{l} \exists x[proved'(x, Marcel')] \\ \exists x[predicted'(x, Marcel')] \\ \exists x[assumed'(x, Marcel')] \end{array} \right\} \tag{2.20}$$

Despite the elegance of treating thematic accents and rhematic accents in a similar way, it is striking that Steedman formalises something that might be described as *contrast* rather than focus. Therefore he misses a correct formalisation for cases with broad focus[25]. He makes the strong semantic claim of an existential presupposition, a view that has been argued against by other researches (Jackendoff, 1972; Rooth, 1985, 1992). Jackendoff (1972) e.g. argues that the pragmatic presupposition cannot trigger an existential presupposition because this would result in a clash for utterances, such as *'Peter introduced* NOBODY *to Sue'*. This utterance obviously cannot presuppose that there is an x that Bill introduced to Sue. He instead proposes the following formalisation, which is related to the concept of *questions-under-discussion* (Roberts, 1996; Ginzburg, 1996):

$$\lambda x[proved'(x, Marcel')] \left\{ \begin{array}{l} \text{is a coherent set} \\ \text{is well-defined} \\ \text{is amenable to discussion} \\ \text{is under discussion} \end{array} \right. \tag{2.21}$$

All these formalisations fail, however, to describe that utterances with identical information structures may have different presuppositions. Example 2.22 shows two utterances with identical information structures, the object *'completeness'* being the rhematic part, *'Marcel proving some x'* (whether existent or not) forming the thematic part. Utterance 2.22a presupposes that Marcel proved something, a presupposition that is not triggered in 2.22b[26] (following Levinson (1983), we use the symbol '>>' to indicate presuppositions). Such a presupposition may be potentially triggered by a very emphatic realisation of the accent only, see section 2.3.3. Alternatively, in a question-context such as *'What did Marcel prove?'*, 2.22a would sound slightly rude, indicating that the questioner should know this fact already. The utterance in 2.22b, on the other hand, would be perfectly felicitous without carrying such an extra-linguistic meaning.

[25]Steedman (personal communication) argues that only narrow focus is kontrast and broad focus is just rheme. But since he treats every accent as contrastive we do not see how he correctly captures broad focus.

[26]Note that most presupposition tests (e.g. negation, question context, modal context) are difficult in the case of focus, since most of them show an *association with focus* (Jackendoff, 1972, p.247).

(2.22) a) It was COMPLETENESS that Marcel proved.
>> Marcel proved something

b) Marcel proved COMPLETENESS
>>?? Marcel proved something

Another implicature from 2.22*a* would be that completeness was the only thing Marcel proved, excluding other — possibly contextually salient — alternatives. The differences between the two utterances in 2.22 can be explained by a notion of *contrast* which we introduce in the next section.

2.3 Defining Contrast

In this section we argue for a concept of "contrast", which is independent from information structural focus. This independence is often doubted and questioned, as the following quotation by Wells shows:

> "[...] it would be of interest to determine whether there is justification, on the basis of its linguistic exponents, for treating 'contrastive' as a distinct category in the system of focus rather, or whether it is in fact indistinguishable from other types of 'new' focus." (Wells, 1988, p. 64)

This does not mean that the general existence of a concept of contrast is questioned. On the contrary, authors are often more inclined to abolish focus and treat everything as contrastive, see Bolinger (1961): "in a broad sense, every semantic peak is contrastive" (p. 87). In other words, accentual highlighting of information is often seen as a sufficient condition for contrast (cf. also Lambrecht, 1994). Molnar (2002) makes the interesting observation that if focusing is always contrastive, "then utterances can only be *used* contrastively and it is only possible to speak of 'clear or less clear instances of contrastiveness'." (p. 150). This presupposes a gradual nature of contrast, which is hardly accepted, though.

If we wanted to draw a simple picture, the linguistic community could be divided into two camps (cf. Umbach, 2004): Semanticists, on the one hand, tend to opt for a uniform treatment of focus and contrast (Jackendoff, 1972; von Stechow, 1990; Rooth, 1992; Krifka, 1993) using the concept of alternative sets. Rooth (1985), for instance, explicitly claims that the general function of focus is to evoke alternatives. Collapsing genuine information structural focus and contrast makes semantic formalisms more elegant because both phenomena can be treated in a concise and uniform way. Such an approach, however, does not distinguish between mere rhythmic accents and semantically meaningful ones; they are all treated as if they were contrastive. The majority of syntacticians and phonologists in this research area, on the other hand, argue for two different kinds of foci, contrastive vs. non-contrastive. They back their claim either with different syntactic properties (Kiss, 1998a) or with phonologically distinct accent types (Pierrehumbert and Hirschberg, 1990).

To be clear from the outset, we follow the view that there **is** a clear difference between information focus and contrast. An illustrating argument for such a division comes from Hungarian: Relying on syntactic properties, the descriptional system of Kiss (1998a) contains a bipartition of focal material into identificational and information foci:

Identificational focus (= contrastive rheme): This kind of focus signals an exclusive reference, i.e. a given proposition only holds for the respective referent and for no others (the English equivalent is the exhaustive focus in it-clefts). A strong empirical support is that identificational foci cannot be combined with universal quantifiers like *'all'* or *'every'*, scalar particles like *'even'* or additive particles like *'also'* (cf. Kiss, 1998a, p. 248). In English also, the utterance *'It was also John that Mary invited to her birthday party'* is ungrammatical. These scalar particles presuppose the presence of alternatives, which, however, are excluded by the exclusive nature of identificational foci.

Information focus (= non-contrastive rheme): Information focus is a purely information structural notion and is defined as the non-presupposed information of an utterance. Every sentence contains an information focus, which can be combined with scalar particles.

In Hungarian, there is a distributional difference between identificational and information foci: "the two types of focus are never optional interpretational variants but are associated with distinct structural positions" (Kiss, 1998a, p. 247). While postverbal accents merely convey non-presupposed information (= *information focus*), preverbal accents are interpreted exhaustively (= *identificational focus*)[27]. Some of her examples are shown in 2.23. Information focus is indicated by small capitals, exclusive identification focus by bold face. Kiss does not further specify the prosodic marking. She notes however that in 2.23*a* not only the information focus but also the particle "be" receive an accent. The constituent representing informational focus remains *in situ* (example 2.23*a*), whereas the constituent representing identificational focus requires a movement to a preverbal position (examples 2.23*b* and *c*). The position of the preverbal accent indicates which of the preverbal constituents is contrasted (compare 2.23*b* and *c*). Note also that the word order of the verb "muttatam" (*introduce*) and the verbal particle "be" (perfect particle) depends on whether it is an utterance with identificational and information foci:

[27] Kiss discusses the data from a generative-syntactic perspective and further supports her claim by specifying different constraints on movement. While identificational foci are moved into a specifier position (which allows them to take scope over other quantifiers), information-foci remain *in situ* and do not enter a scope relation (cf. Kiss, 1998a, p. 253ff).

(2.23) a) *Tegnap este be mutattam Pétert* MARINAK.
last night PERF introduce Peter Mary
'Last night I introduced Peter to MARY'.

b) *Tegnap este* **Marinak** *mutattam be Pétert.*
last night Mary.DAT introduce PERF Peter.ACC
'It was **to Mary** that I introduced Peter last night.'

c) **Tegnap** *este mutattam be Pétert Marinak.*
last night introduce PERF Peter Mary
It was **last** night that I introduced Peter to Mary.

Kiss (1998b) criticises that these obvious differences between identificational and information focus are not acknowledged in current semantic formalisms. The structured meaning theory of focus, e.g., would assign the structure in 2.24 to sentences 2.23*a* and *b* alike (cf. Kiss, 1998a, p. 247)[28]:

(2.24) ASSERT ($< \lambda$x.introduced (I, Peter, x), Mary $>$)

The influence of contrast on word order and prosody in Hungarian strongly suggests that contrast is important for language understanding.

We assume that what has often hampered mutual understanding in this area is that the term "contrast" has been used to denote a variety of different concepts, many of which are only loosely related. The terminological problem, which is partly caused by influences from different linguistic areas, is nicely summarised in the following quotation:

> "Contrast is one of those general conceptions we all are acquainted with, everyone having particular experiences and images correlated with it. Contrast is absolutely not a linguistic term." (Rudolph, 1996, p. 3)

However, contrast has long been dealt with purely linguistically, especially syntactically. Since Chomsky and Halle (1968) described accent placement that deviates from a syntactically defined default-position with the term "contrastive stress", many researchers set out to define structural rules to identify this default-position (termed "normal stress") and — in turn — to define contrast (Chomsky and Halle, 1968; Gussenhoven, 1983; Selkirk, 1984, 1995). The idea that there exists some sort of "normal stress", which can be defined on a purely structural basis, was criticised by various authors (especially Bolinger, 1961; Schmerling, 1976), presenting counter-examples with an identical syntactic structure but a different accent placement. Bolinger showed that the semantic weight of the constituents influenced the position of the nuclear accent (see example 2.25), whereas Schmerling discussed the importance of the (situational) context for accent placement, see example 2.26.

[28] For a better understanding the semantic form is simplified, ignoring the temporal adverb.

(2.25) I have a POINT to make.
vs. I have a point to EMPHASISE.

(2.26) Truman's DIED.
vs. JOHNSON's died.

Bolinger's example (2.25) shows that the accent may fall on the verb if this is semantically heavy — although this is not what syntactic rules predict. In Schmerling's example (2.26), the accent position is dependent on the circumstances at the time of utterance: The former American president Truman had been hospitalised in serious condition and according to Schmerling daily reports were given concerning his critical condition. It seemed rather unlikely that he would survive this crisis. As Truman's health condition was in everybody's mind, only the new fact — that he had died — needed accenting. Johnson's health condition, on the other hand, also had been in the news for some time but he was apparently recovering from a heart attack and his health condition was not that newsworthy anymore. Therefore, accenting Johnson was necessary.

To reconcile the criticism brought forward against the notions of normal and contrastive stress with the syntactic regularities observed, Ladd (1980) proposed the notions *broad* and *narrow* focus. These are not just different labels for the same concepts but are to be interpreted as a continuum rather than two categories, i.e. a focus can be arbitrarily broad or narrow. Generally, narrow focus does not allow for focus projection, i.e. the focus is associated with the word that bears the accent only. Broad focus, on the other hand, allows for the interpretation of focus domains that exceed the domain of the actual accented word. Focus domains for utterances with broad focus are hence said to be ambiguous, compare example 2.27*a* with a narrow focus on the colour adjective to 2.27*b* with a broad focus.

(2.27) a) He was warned to look out for an exconvict with a [$_F$ RED] shirt.
b) He was [$_F$ warned [$_F$ to look out for [$_F$ an [$_F$ exconvict [$_F$ with [$_F$ a red SHIRT.]]]]]].

In the narrow focus sentence (2.27*a*), only the adjective is focused. In 2.27*b*, constituents of different sizes may be in focus, such as the *'shirt'* only or the *'exconvict with the red shirt'*. Intuitively, it is desirable to account for the difference in perception apparent in the examples in 2.27. While the utterance with the narrow focus (2.27*a*) triggers a contrast to other potential exconvicts with a differently coloured shirt and might be interpreted even as corrective, such effects are not directly possible with the broad focus in 2.27*b*. In section 2.3.1, we describe accounts in which alternatives are seen as the key to contrast. In section 2.3.2, we then explain why it is so difficult to define contrast uniformly. One reason is that contrast can e.g. hold between concrete elements in the discourse universe but also between the mere assumption about something and the proposition of the current utterance. In section 2.3.3 we finally discuss previous work on the prosodic marking of contrast.

2.3.1 Alternatives

Intuitively, it is the presence of alternatives that should distinguish contrast from information structural focus: While contrast can be used to exclude or to trigger alternatives, information structural foci merely provide new information[29]. This difference between contrast and focus has entered linguistic theory in many different notions, such as *i-focus* vs. *c-focus* (Zubizarreta, 1998), *information focus* vs. *identification focus*[30] (Kiss, 1998b), and — in a related sense — *narrow focus* vs. *broad focus* (Ladd, 1980). Accordingly, focus is formalised as the instantiation of a variable (e.g. x) in an open-proposition (e.g. λ x[proved'(x,Marcel')]). However, in pure logical terms, any focal instantiation in a given context is *a priorily* selected from an infinite set of possible alternatives. Some authors therefore proposed two different sets of alternatives, an open, infinite one and a closed set containing only a limited number of candidates (Halliday, 1967; Chafe, 1976; Rooth, 1992). Chafe, for instance, argued that "contrastive sentences are qualitatively different from those which simply supply new information from an unlimited set of possibilities" (p. 34).

In this section we first concentrate on speech production and perception. In the speech production process, the presence (and salience) of possible alternatives may influence speech planning (see discussion on microplanning in Levelt, 1993, p. 149). Note, however, that a contrast between referents or actions may not be salient to the hearer in the same way as it is to the speaker. It is only from the actual realisation of the utterance that the hearer might be able to grasp that there was some kind of contrast for the speaker (although the listener might not be able to identify the items to which the contrast was established).

We shall first discuss how the presence of possible alternatives influences production (section 2.3.1). In section 2.3.1 we then discuss the power of contrastive utterances to implicate the presence of possible alternatives. It has been noticed that certain constructions (or certain prosodic realisations) seem to have two interpretations: First, the literal, compositionally derivable meaning of the utterance and second, a presupposition concerning possible alternatives. In focusing on production and perception, we can isolate two different impacts of contrast, the exclusion and triggering of contrast:

1. Contrast *production*: possible alternatives are *excluded*, i.e. the given proposition only holds for the respective item and not for other potentially salient alternatives (cf. Bartels and Kingston, 1994; Kiss, 1998b).

   "The concept of *contrastiveness* can be characterised as 'assertion of a propo-

[29]Some authors, however, make a distinction between contrast and alternatives: Lee (1999), e.g., associates focused elements with alternatives but not with contrast: "A genuine Focus, I would argue, has no notion of contrast except that of 'alternatives' [...] The focused element alone is highlighted and other alternatives are simply shadowed and ignored or excluded at the moment of utterance, even if 'evoked' in explicit linguistic context." (p. 326f).

[30]Note that Kiss further divides identificational focus into contrastive and non-contrastive uses.

sition accompanied by explicit denial of salient alternatives'." (Bartels and Kingston, 1994, p. 1)

2. Contrast *perception*: possible alternatives — with different, possibly underspecified propositions — are *triggered* (Krahmer and Swerts, 2001).

The Production Perspective

Kiss used Hungarian word order differences to make the case for a differentiation between focus and contrast. In the languages of the world, we find a large variety of contrast-markers, which should be a further argument for its independent impact. Languages like English, which have less word order freedom than e.g Hungarian, often make use of special syntactic operations, such as it-clefts or pseudo-clefts to mark an item as contrastive. Another possibility is the use of restrictive particles (such as *only*), which are present in one form or another in all languages to mark identificational focus. Example 2.28 shows some English and German utterances with restrictive particles. Note that for both the Hungarian example (2.23) and the German and English examples (2.28) interpretation strongly depends on the position of the nuclear accent.

(2.28) b) *John hat nur seiner* TOCHTER *ein neues Fahrrad geschenkt.*
John has only his daughter a new bicycle given
John only gave his DAUGHTER a new bicycle.

b) *John hat nur* SEINER *Tochter ein neues Fahrrad geschenkt.*
John has only his daughter a new bicycle given
John only gave HIS daughter a new bicycle.

However convincing the presence of alternatives is for describing contrast, we should be careful not to overestimate their influence. Example 2.29*a* instantiates the concept of favourite cities with three cities, Geneva, Kyoto, and London. Let this be the context, in which the utterances 2.29*b* and *c* are to be interpreted:

(2.29) a) My favourite cities are Geneva, Kyoto, and London.
my-favourite-cities = {Geneva, Kyoto, London}

b) I will spend my holidays in LONDON.
London $\in$ my-favourite-cities

c) I will spend my holidays in PHILADELPHIA.
Philadelphia $\notin$ my-favourite-cities

When the respective utterances are produced with an accent on the respective city only, 2.29*b* does not seem very contrastive, whereas 2.29*c* surely does. A simple diagnostic test is to insert an adversative particle, such as *'however'*, *'though'* or *'but'* and evaluate whether the utterance becomes inappropriate in a given context: *'However, I will spend my holidays in London'* seems unusual, whereas *'However, I will spend my holidays in Philadelphia'* is suitable. This is surprising, given that the utterance in 2.29*b* chooses one element from the set of favourite cities, whereas 2.29*c* does not. Why is that so? One possible explanation is that the utterance in 2.29*c* establishes a contrast to the *implicit* assumption that people visit their favourite cities during their holidays. This different application of contrast is discussed in more detail in section 2.3.2. If we replace the word *'holidays'* by *'summer holidays'* with a rising accent on *'summer'*, both utterances in 2.29 trigger the implicature that the *'winter'* holidays are spent in another city. The utterance is then perceived as having a double contrast, one between *'summer'* and *'winter'* holidays and one between the travel destinations in these periods of the year (although implicit in one case). We shall return to these cases with a double contrast later.

Excursus: The "Border" to Corrections

Phoneticians and phonologists often use the following elicitation criteria for informational and contrastive foci: Information foci are triggered in a question-context, which focuses on a specific word, while contrastive foci are almost exclusively brought about by a correction context. Example 2.30 — taken from the English translation of Frota (2002), p. 319 — shows the benefits of such a definition: the answer to both questions is both lexically and structurally identical and hence only differs in the pragmatic context (Q_1 vs. Q_2).

(2.30) Q_1: Who gave spices to the journalists?

Q_2: Did the Mozambicans give spices to the journalists?

A: [$_F$ The Angolans] gave spices to the journalists.

Note that there are several difficulties associated with question contexts of the form Q_1: First, the non-elliptic answer is rather unnatural and results in a fairly long stretch of unaccented post-focal material, which might have an influence on the prosodic realisation of the rhematic accent. Second, the prosody of the answer will be influenced by the prosodic structure of the question Q_1. Q_1 can signal much more than a simple information request, such as bad understanding (*'Excuse me,* WHO *did it?'*) or disbelief (*'That can't be true, that the Angolans gave spices to the journalists; tell me again:* WHO *did that?'*), etc.

Corrections are closely related to the exclusion of possible alternatives, because they explicitly deny that a certain proposition holds for an alternative item. Nevertheless, the explicit correction differs from an exhaustive identification in both a conceptual and a realisational way. Conceptually, corrections are some sort of meta-linguistic contrast (von Heusinger, 1999; Krahmer and Swerts, 2001), which pursue a change of the interlocutor's belief state. This is obviously different from a denotational contrast (notion used by Lee

(1999)) that merely expresses that two items have different properties. This conceptual difference manifests itself in different ways, which will be the focus of the present excursus.

We want to discuss examples from Italian and Catalan on the one hand and data from English and German on the other hand to show different peculiarities of corrections. While in the former group of languages, corrections directly influence the grammar of the language, its impact on the latter is merely gradual. Ladd (1996) described Italian as a language that resists deaccenting of given material. This is exemplified in 2.31*a* (taken from Ladd (1996), p. 176). The English translation, which preserves the prominence structure of the Italian original, sounds very odd, at least for speakers of German or English because the accent falls in both clauses on "posto" (*'place'*). We would possibly prefer an accentual structure that highlights the contrasts (*'The investigations are helping to put* BACK *in order things that have got* OUT *of order'*). What makes languages such as Italian particularly interesting for the present claim is that in Italian deaccentuation becomes possible in the case of a correction, see example 2.31*b* (also taken from Ladd, 1996, p. 177):

(2.31) a) *[le inchieste] servono a mettere a* POSTO *cose andate fuori*
[the investigation] serve to put to place things gone out of
POSTO.
place
'The investigations are helping to put back in ORDER things that have got out of ORDER'.

b) *Non ho detto* CASA *bianca, ho detto* COSA *bianca.*
not past say house white, past say thing white.
'I didn't say white HOUSE, I said white THING.'

Vallduví and Engdahl (1996) report a related case in Catalan, in which the syntactic structure of corrections may well deviate from ordinary syntactic rules. They describe Catalan as a language with a fixed accent position, in which the accent has to be realised on the last element of the core clause[31]. They give the example in 2.32 to show that a structure that is possible in a corrective context (2.32*a*) is ungrammatical otherwise (2.32):

(2.32) a) *Posarem "*EXTENSIÓ*" en aquesta oració (no "*INTENSIÓ*")*
1p-fut-put extension in this sentence not intension
'We'll use "EXtension" in this sentence (not 'INtension').'

b) Q: Which term will we use in this sentence?
A: *Posarem "EXTENSIO" en aquesta oració.

A different kind of support is entirely phonetic and has surfaced mainly in speech recognition applications and research. Dialogues of the form 2.33 are frequent in human-computer

[31] To encode different information structures, non-focal elements have to be moved out of the core clause (by fronting or right-dislocation).

interaction, and it has been noticed that recognition errors are two to three times higher in the corrections than in the initial input (Levow, 1998).

(2.33) System: Where do you want to go?
User: I want to go to POZNAN.
System: When do you want to go to Boston?
User: No, not to Boston; I want to go to POZNAN!

Therefore, the phonetic and prosodic properties of corrections have been intensively investigated in human-machine interaction (Oviatt et al., 1998a,b). The main finding is that corrections are fairly hyper-articulated, i.e. they have longer durations, a higher amplitude and more extended f_0-movements. This information has been successfully employed to train decision trees, which obtained a fairly high accuracy of 75% (Levow, 1998).

While the English example merely exhibits a gradual difference between corrections and foci, corrections reportedly influence the grammatical structure in languages such as Italian and Catalan. That is, they are not merely an exaggerated form of contrast (as could be inferred from the English example) but are structurally different. This should help to refrain from using corrective contexts when investigating contrast. Unfortunately, these two concepts are often confused, especially in experimental phonetic research. Analysing corrections instead of contrast has been favoured by the experimental procedure: While it is straightforward to trigger corrections, elicitation of contrast is more difficult, especially in utterances with a contrastive rheme only.

The Perception Perspective

Krahmer and Swerts (1999) argue that contrastive accents trigger a presupposition about alternatives. In their model, an utterance such as '*BLUE square*' (with an accent on the adjective) indicates the presence of an item x, which is blue and has the shape of a square, but it also presupposes the existence of an element y, which is a square but not blue. In any other context, the utterance '*BLUE square*' seems inappropriate. Krahmer and Swerts (1999) experimentally showed that subjects were able to infer the appropriate pre-contexts.

Watson et al. (2004a,b) experimentally tested the interpretation of the accent types H* and L+H*. They used an action-based version of the visual world paradigm (Tanenhaus et al., 1995). Participants were seated in front of a computer screen that depicted four movable objects. They were instructed to click on some of them or to move them. Watson and colleagues made use of "cohorts", i.e. words with identical onsets (e.g. '*camel*' and '*candle*'). The target word either belonged to the contrast set (e.g. animals, defined by two previous instructions to click on a camel or to move a dog) or not (e.g. a candle). Eye-movements were monitored when the target word was produced with an H* accent or an L+H*. For L+H* they found significantly more looks to the contrastive referent than to the non-contrastive one. For H*, however, there was no such effect. These results indicate that an L+H* accent is a possibility to mark an element as contrastive, but that even when it is not marked as contrastive it can be perceived as contrastive.

Even more interesting than utterances with a contrastive rheme only are utterances with both a contrastive theme and rheme. It is often claimed that utterances with a contrastive theme and rheme (a.k.a. contrastive topic) are realised with a particular intonation pattern. Broadly speaking, this consists of an accentual rise on the contrastive theme and an accentual fall on the contrastive rheme. The pitch stays high between these accents. Notationally, we use a forward slash before the word that contains the accentual rise; the word with the rising accent is highlighted with bold face. The pitch fall is signalled by small capitals and a backward slash following the accented word[32]. An example is given in 2.34. If the answer is uttered neutrally, i.e. with a pitch accent on "zuhause" (*'home'*) only, it asserts nothing more than the literal meaning, namely that the speaker was at home (see utterance a) in Figure 2.4). If, however, the answer is produced with a hat pattern (see utterance b) in Figure 2.4) it rather implies that somebody else might have been at the site of crime, which might draw the investigators' attention to other possible suspects.

(2.34) Q: Where were you at the time of the murder?
A: /**Ich** war ZUHAUSE\\.
(Engl: I was at home.)

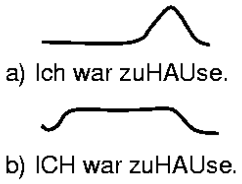


Figure 2.4: The utterance "Ich war zuhause" (*'I was at home'*) spoken neutrally (a) and with a hat pattern (b).

There has been a considerable amount of research into the semantics of utterances with hat patterns in German (Jacobs, 1996; Büring, 1997; Steube, 2001; Mehlhorn, 2001, among others). In utterances with two quantifiers or one quantifier and a negation particle, the scope of the quantifiers is reversed when the utterance is produced with a hat pattern, compare examples 2.35*a* and *b*:

(2.35) a) ALLE Politiker sind nicht korrupt.
$\forall x$[Politiker(x)$\rightarrow \neg$korrupt(x)]
'All politicians are not corrupt.'

[32]A more detailed phonetic description is found in section 2.3.3.

b) /**Alle** Politiker sind NICHT\ korrupt.
$\neg\forall x$[Politiker(x)$\rightarrow$ korrupt(x)] $\equiv \exists$x[Politiker(x) $\wedge\neg$ korrupt(x)]
'Not all politicians are corrupt.'

Mehlhorn (2001) argues that contrastive themes (I-Topics) establish an implicit contrast that the addressee has to infer. Büring (1997), who studied the uses of hat patterns fairly thoroughly, also discusses *partial topics*, i.e. when a given response does not completely answer the question. Another way of usage is to implicate something about other salient topics, as shown in example 2.34. This sort of tune can also be used to evade a question and to implicitly change the topic of the discourse.

From a formalisation perspective, both the exclusion of alternatives as well as their triggering is not much different. In the literature, there have been various approaches to formalising them: Karttunen (1974), for instance, discussed it under the term *context extension*, and Lewis (1979) defined a *rule of accommodation*, which is cited below:

> " if at time t something is said that requires presupposition P to be acceptable, and if P is not presupposed just before t, then – *ceteris paribus* and within certain limits – presupposition P comes into existence at t." (Lewis, 1979, p. 340)

Accommodation, which was described as a "fundamental mechanism in pragmatics" by Thomason (1990), p. 333, makes perception and interpretation very flexible. Presuppositions are very common in communication; it may be speculated that it is them that make communication so efficient, but it may also be them that cause so many misunderstandings. It may happen quite easily that intended "neutral" utterances are interpreted as contrastive or even corrective.

This flexibility of perception and interpretation makes them a rather inappropriate diagnostics for identifying contrast. In the next section we shall discuss further possible definitions of contrast and propose an operational working definition.

2.3.2 A Definition Problem

Contrast in general, its realisation, interpretation and formalisation have been investigated in many linguistic subfields, such as phonetics, phonology, syntax, semantics, and pragmatics. However, the different disciplines have developed quite different conceptions when referring to *contrast*, which has hampered scientific progress. Semanticists define contrast with respect to alternatives, but as shown in 2.3.1, the presence of alternatives might be a necessary condition for contrast but is definitely not a sufficient one. Furthermore, we have shown examples in which contrast is not established between two concrete referents but between assumed propositions and actually uttered propositions (cf. example 2.29). This very broad sense of contrast is used by few researchers only (cf. Hedberg, 2002). Most phoneticians mainly investigate corrections, which we described are a very special sort of contrast. Here, we are not able to provide a uniform definition of contrast but we want to draw attention to different sorts of contrast. Furthermore, we shall describe some of the problems associated with defining contrast and propose a working definition.

We first discuss some examples with a contrastive discourse relation because these cases are mostly clear-cut. In rhetorical structure theory, RST (Mann and Thompson, 1988), contrast is defined as a multinuclear rhetorical relation, which reflects the presence of two equipollent parts. Contrast may hold within two clauses of a sentence or – on a higher level – between two larger segments of a text. In many languages, intra-sentential contrast is signalled by contrastive conjunctions. Often, languages have more than one adversative conjunction that can signal a contrastive relation between two clauses, e.g. "während" (*'while'*), "aber" (*'but'*), "obwohl" (*'although'*), "wohingegen" (*'whereas'*):

(2.36) a) I like coffee but I hate tea.

b) While he was asleep, I was reading.

c) I am playing soccer, although I am not very good at it.

Although each of the utterances in 2.36 exhibits a contrast between its two clauses[33], the contrasted elements differ. In 2.36*a*, only the utterance's predication, the verb phrase, is contrasted: {<like'(coffee',I'), hate'(tea',I')>}. The topic is identical in the two clauses and only the rhematic parts are contrasted. In 2.36*b* the contrast spans the two propositions {< asleep'(he'), reading'(I')>}. That is, there is a contrast in both the theme and the rheme. Example 2.36*c* is different from the ones just discussed because there, the contrast is expressed between an assumed proposition and an actually stated one: From its first clause we might draw the inference that people who play soccer are good at it; so the second clause contrasts this implicit assumption. It is interesting to note that even if we had not drawn this inference after hearing or reading the first clause, we would have drawn it after having processed the adversative conjunction and the second clause.

[33] Intuitively, the conjunctions in examples 2.36*a–c* exhibit a different "degree" or "amount" of contrast. The conjunction "aber" (*'but'*) is stronger than the other conjunctions that also allow for a temporal (*'while'*) or concessive interpretation (*'although'*). It is an open question whether the contrast is mainly established by the lexical meaning of the respective conjunction or by the propositions that are contrasted. All these conjunctions add some further meaning. For a better understanding we discuss the difference in meaning, which results from the use of *'but'* and *'and'* in example 2.37 (taken from Bach, 1999):

(2.37) a) Shaq is huge but agile.

b) Shaq is huge and agile.

In terms of truth-conditional meaning, both utterances make the same predications, i.e. the assertions of both utterances are only true if Shaq is both huge and agile. Nevertheless, 2.37*a* adds the meaning (presupposes, implicates) that it is contrary-to-expectation for a huge person to be agile. This additional meaning is absent in example 2.37*b*.

In some languages, this intuitive difference between 2.36*a* and *b* on the one hand and 2.36*c* on the other hand is differentiated linguistically. In Polish, e.g., two different kinds of contrastive conjunctions are used, depending on whether the contrast is established between items explicitly encoded in the utterance or between an utterance and an implicit assumption: "natomiast" and "ale"[34]. The contrastive conjunction "ale" (cf. example 2.38*a*) is paramountly used in cases in which a possible implicature (that museums are interesting) is contrasted (cf. example 2.36*a*). The contrastive conjunction "natomiast" in examples 2.38*b*-*d*, however, is used when there is a contrast between two explicitly mentioned items or predications[35]. Example 2.38*c* shows that the use of "natomiast" is not restricted to referential expressions in a strict sense, but is also possible when temporal phrases are contrasted. Example 2.38*d* further shows that the contrasted clauses need not be syntactically parallel, a rule that could be inferred from the previous examples. *Semantic* parallelism between the two clauses is sufficient.

(2.38) a)

To	*jest*	*muzeum,*	**ale**	*nie*	*jest*	*zbyt*	*ciekawe.*
This	is	museum	but	not	is	overly	interesting

This is a museum but it's not very interesting

b)

To	*jest*	*muzeum*	**natomiast**	*to*	*jest*	*szpital*
This	is	museum	and	this	is	hospital

This is a museum and that is a hospital.

c)

W	*poniedziałek*	*pracuje*	*w*	*szpitalu*	**natomiast**	*we*	*wtorek*	*w*	*bibliotece.*
On	Monday	work$_{1st-sg}$	in	hospital	but	on	Tuesday	in	library

On Monday I work in the hospital and on Tuesday I work in the library.

d)

Andrzejowi	*się*	*podabał*	*film*	**natomiast**	*Anna*	*usneła.*
Andrew$_{dat}$	REFL	liked	movie	but	Anna	fell asleep

Andrew liked the film but Anna fell asleep.

The examples that exhibited a contrast in both theme and rheme (2.36*b*, 2.38*c* and *d*) show that contrast is possible in both information structural primitives, as formalised by Steedman (2000). These examples only present counter-evidence for the position that contrast is synonymous to focus only[36]. The fact that Polish uses two different conjunctions clearly shows that contrast is no monolithic concept. On the other hand does this

[34] Note that the conjunction "natomiast" is fairly formal. A more informal and more commonly used conjunction is "a", which is, however, semantically too underspecified to be used in the present argument.

[35] As Ladd (personal communication) pointed out, there is a similar distinction in Rumanian between the contrastive particles "dar" and "iar".

[36] There are researchers, however, for whom all contrast becomes focal (von Fintel, 1994; Steube, 2001; van Hoof, 2003). Another way of formalising this issue is to treat contrast as an additional information structural primitive (Molnar, 2002).

peculiarity allow us to propose a narrow working definition of contrast, which only includes such contrastive theme–rheme pairs as shown in 2.38*b*–*d*. In this narrow sense of the term, contrast is established between two propositions about comparable referents, locations or times, which can be referred to as *semantic parallelism*. This is in line with other proposals by Prevost (1995) and Pulman (1997) as well as with experimental evidence provided by Theune (1999). Semantic parallelism is not restricted to intra-sentential contrast but may span larger contexts as shown in example 2.39.

(2.39) I was raised in a large family. I have two brothers and a sister. My brothers, who are younger than we girls, still live at home, whereas my sister and I are already married. I have one child. My husband and I live in the suburbs of London. My sister has three children. I suspect she really liked being in a large family.

2.3.3 Prosodic Issues

In this section we review some of the studies that have investigated the prosodic realisation of contrast. Broadly speaking, they can be grouped into those that investigated the realisation of contrastive *rhemes* and those that were mainly interested in utterances with both a contrastive theme *and* a contrastive rheme.

Contrastive Rheme Accents

There have been various studies that explored the issue of whether contrastive rhemes are prosodically marked in a different way than non-contrastive ones. However, it is difficult to compare their results because the granularity of analysis ranges from mere introspection over phonological analysis to fine acoustic measurements. Therefore, there are different views on whether there is a prosodic difference between contrastive and non-contrastive accents (in the same position of an utterance). Researchers interested in intonational meaning tend to claim phonological differences between neutral and contrastive foci. Their conclusions are often based on the analysis of only a few prototypical examples. Pierrehumbert and Hirschberg (1990) thus found for English that contrastive accents are marked with an L+H*, whereas non-contrastive ones are marked by an H*. They claim that a speaker uses an L+H* accent to "convey that the accented item — and not some alternative related item — should be mutually believed" (p. 296f). However, they do not indicate whether this kind of contrast-marking is applicable to both themes and rhemes (and the examples they give are entirely rhematic). Hedberg (2002) found in a corpus study that L+H* was associated with topics in 58% of the cases and with foci in 40% of the cases[37] (summing up contrastive and non-contrastive uses). As regards contrastiveness, L+H* was used to convey (topical or focal) contrast in 57% of the cases.

[37] Interestingly, L+H* was used by Steedman (2000) for theme accents only.

Here we shall mainly concentrate on a selection of results gained from *experimental* prosodic studies. Often contrastive rhemes are elicited in a *corrective* context, which may exhibit hyper-articulated segmental and prosodic features. We shall see that there is a large variation even in this comparatively small area of research, ranging from categorical differences in accent type to gradual variation (mainly in peak height). There are obvious cross-linguistic differences in the prosodic marking of focus, which are related in a complex way to the syntax and morphology of focus marking (cf. Frota, 2002). For Romanic and Slavic languages, researchers often report categorically distinct accent types, while for Germanic languages, there merely seem to be gradual differences.

Categorical differences. Oliver and Andreeva (to appear) analysed peak height and peak alignment in different focus structures in Polish and Bulgarian. Identical utterances were read in different focal conditions with two different speech rates, normal and fast. We are concentrating here on their results of narrow foci (elicited by a corresponding wh-question) and contrastive foci (elicited by a correction). For Polish, they found a significant difference in the peak alignment of these two accents, with non-contrastive ones being aligned earlier than contrastive ones. In ToBI-notation, this might be described as H*+L (or H*) for non-contrastive narrow foci and L*+H for contrastive foci. Bulgarian similarly exhibited a difference in accent type, H* for non-contrastive foci and H*+L for contrastive ones.

Selkirk (2002) explored whether contrastive and representational (informational) focus are to be distinguished in phonological or intermediate syntactic representation. Because she assumes a neutralisation of the presumed difference in utterance final position (Ladd, 1996), she investigates their prosodic characteristics in foci that precede another informational focus. To avoid any influence from an upcoming prosodic boundary, she carefully controlled the right context. However, the target utterances were no "minimal pairs", i.e. identical utterances in pragmatically different contexts. Instead, she investigated different sentence types (object–verb–subject structures vs. right node rising), in which only the syntactic structure from the target word onwards was identical. One of her target pairs is shown in 2.40*a* and *b*, with the target word underlined. The utterance in 2.40*a* is an utterance with a simple syntactic structure, subject–verb–object. For the utterance containing the contrastive focus (2.40*b*), Selkirk made use of a special syntactic structure, called right-node-rising (RNR) to use the same right context as in example 2.40*a*. That is, the (unexpressed) object of the verb *'remaster'* is the same as the object of the target verb *'remanipulate'*.

(2.40) a) We've been <u>remanipulating</u> the recordings of the Black Cat label.

b) We've managed to remaster without <u>remanipulating</u> the recordings of the Black Cat label.

Selkirk found a phonological difference between contrastive and informational foci: contrastive FOCUS has a L+H* pitch accent followed by a L- phrase accent, whereas informational focus (in Selkirk's terms presentational focus) is realised by a H* pitch accent without a phonological phrase break. It is doubtful, however, whether these prosodic differences can be attributed to the difference in contrast only. It rather seems that other confounding factors, such as ellipsis, the different complexities of the compared syntactic structures and the different sentence lengths might have contributed to the alleged phonological difference between informational and representational foci.

Face (2002) compared the realisation of corrective and neutral foci in words in initial and medial position (i.e. words that are followed by at least one other content word) using 20 speakers. Interestingly, he reported different strategies for marking corrective foci in Spanish: either an L*+H with a higher f_0-peak compared to broad focus realisations, an H- or L- phrase accent or an L+H* pitch accent. Especially the H- phrase break is reportedly used for broad focus cases as well but Face claims a significant distributional difference with more H- phrase breaks in corrective foci. He acknowledged that it was still an open question whether these different kinds of rhematic contrast-marking were perceptually equally important. Based on this results he considers a continuum in the degree of focus marking:

> "In a similar manner, it will be important to investigate whether these acoustic patterns are perceived as focal in a categorical manner or whether there is a continuum of degree of focus which is perceived based on a particular parameter (e.g. the location of the peak) or by use of different intonational strategies (e.g. L+H* vs. L-)." (Face, 2002, p. 91)

Continuous differences. For German, English, and Dutch, researchers claim no or at least continuous differences between contrastive and non-contrastive rheme accents. Ladd (1983) e.g. claims that contrastive foci exhibit a higher degree of emphasis, i.e. they are extra high or boosted. Similarly, Bartels and Kingston (1994) report that contrastive rhemes exhibit a higher peak together with a steeper slope.

Contrastive Theme and Rheme Accents

More relevant for the present thesis are experimental studies on contrastive theme *and* rheme accents. The notion of contrast in these studies is often better defined than the rhematic contrast only, as discussed in section 2.3.2. In the literature, different notions for *contrastive themes* (Steedman, 2000) have been used, such as *S-Topic* (Büring, 1997), *I-Topic* (Steube, 2001; Mehlhorn, 2001), and *Topic-Focus* (as opposed to *Focus-Focus*, see Kadmon, 2001). We prefer the term contrastive themes because it does not have all the theoretical ballast topics have, such as referentiality and sentence-initiality (see section 2.2.1). Intuitively, we postulate that — unlike contrastive rhemes — contrastive themes cannot appear in isolation, i.e. a contrastive theme always implies a contrastive rheme (cf. also Alexopoulou, 1996; van Hoof, 2003)[38].

There have been relatively few experimental studies that compared the realisation of contrastive and non-contrastive themes. Most semantic formalisms refer to impressionistically defined intonation contours, and little is known about the actual prosodic properties. Contrastive theme-rheme pairs are generally associated with a hat pattern, while non-contrastive themes lack this particular intonation contour. The term hat pattern was initially coined by Cohen and t'Hart (1967) to describe a contour with an initial rise, a sustained high pitch, which finally falls onto the nuclear accent. This contour was described as the basic intonation contour for Dutch. Meanwhile, there have been various notions for this particular pattern, with intonational descriptions as varied as the terms. The most commonly used terms in German articles are "Hutkontur" (*'hat pattern'*), used e.g. by Mehlhorn (2001) and Steube (2001), "Brückenakzent" (*'bridge accent'*), employed by Wunderlich (1991) and Büring (1997), and "Wurzelkontour" (*'root contour'*), originally used by Jacobs. A more functional notion often found is "I-Topicalisation" Jacobs (1997), referring to the fact that certain constituents can only be topicalised to the preverbal field (German "Vorfeld") if this is licensed by a particular intonational form (hence "I" for intonational). Even for German, there is no concise description of this frequent contour. Researchers differ both in the choice of tone for the first and the last accent, as well as for the semantic interpretation:

- Wunderlich (1988) describes the *bridge accents* in terms of the following tone sequence: H* H L*. The H without diacritics signals a floating tone, which is associated with all unspecified intervening syllables that are not otherwise associated. According to Wunderlich, the bridge accents can be used in a large variety of cases, such as broad focus, proverbs, contrast, alternative questions, lists, to signal different syntactic bracketing, syntactic extractions and gapping (see Wunderlich, 1991, p. 242f)

[38] This view is not in line with Steube (2001), however, who claims that such utterances only evoke alternatives to the theme, while a contrast to the rhematic part is not assumed (p. 20). Furthermore, Steedman (2000), p. 661, discusses utterances with contrastive themes only as shown in example 2.41:

(2.41) Q: Does Marcel love opera?
A: Marcel likes MUSICALS.
L+H* L+H%

He argues that the answer offers a new theme without stating a rheme at all. According to Steedman this implicates a lack of confidence regarding the relevance of the answer in light of the question. If there were an H* LL% contour on *'musicals'*, this would be interpreted as either denial or affirmation, depending on whether liking operas implies loving musicals or not. In German, the answer is likely to be produced with a hat pattern: /**MU**sicals MAG er, which is rather to be interpreted as a contrastive theme–rheme pair.

- Féry (1993) distinguishes between two different kinds of hat patterns. If we follow her argument correctly, *hat pattern 1* (which she describes as instable) is signalled by a sequence of H* and H*L. The genuine hat pattern seems to be represented by *hat pattern 2* with a L*H H*L contour (p. 149ff). She discusses the problem that the difference between these two contours "is not always phonologically clear-cut." (p. 151). This is probably the reason that she does not describe semantic differences between these contours.

In recent linguistic publications there have been proposals that the thematic (rising) accent has to fulfil certain prosodic constraints in order to license a contrastive topic accent. Jacobs (Jacobs, 1997), e.g. discusses that it is not the pitch rise *per se* that signals a contrastive topic but rather the fact that the rise is preceded by a noticeable trough (which is why he prefers the notion root contour).

Mehlhorn (2001) was the first to analyse the acoustic phonetic differences between the rise in a hat pattern and other prenuclear rises (or — in Féry's terms — the difference between a rise in *hat pattern 1* and *2*), in more detail. She investigated the production and perception of topical elements in contrastive and non-contrastive contexts and found that topics in contrastive contexts exhibited different phonetic properties from topics in non-contrastive contexts. The main differences were a greater f_0-excursion of the prenuclear rising accent (theme) as well as longer durations for contrastive topics.

In the present thesis we are mainly interested in the prosodic properties of themes in contrastive and non-contrastive contexts. Intuitively, prenuclear rises in contrastive contexts seem to start from a lower point than those used non-contrastively. To be able to compare two versions of prenuclear accents, it is important to use "minimal pair" utterances, i.e. identical utterances, which only differ in the pragmatic context they are produced in. This forces us to investigate unmarked sentence structures that are not *a priorily* interpreted as contrastive.

Summary

In this section we first presented an overview over intonational annotation schemes and approaches to model intonational meaning. Since intonational meaning can be most generally captured in information structural terms, we briefly reviewed some early information structural theories that assume two information structural primitives and described the need for a finer distinction, which is met in modern hybrid theories (Vallduví and Engdahl, 1996; Steedman, 2000). In Steedman's approach, which is more general than the one by Vallduví and Engdahl, prominent information is acknowledged in both the *theme*, i.e. the part of the utterance that connects the utterance to the context and in the *rhemes*, which indicates the new contribution in a given context. The examples Steedman uses are those that we describe as contrastive theme-rheme pairs.

As the notion of contrast is no monolithic concept, we spent some time in disentangling the different uses found in the literature. We discussed that it is mainly the extra-linguistic nature of contrast that makes it difficult to provide a uniform definition of contrast. Some researchers adopt a very wide definition of contrast, as found in rhetorical structure theory. The other extreme is to equate corrections with contrast. This we objected because in some languages, corrections may alter the syntactic structure. We showed that, on theoretical grounds, it is difficult to distinguish rhematic contrast from information focus if we only use the notion of alternative sets. This does not mean that there are no sentences with a rhematic contrast only. However, at present there are no appropriate means to operationally define them. In the remainder of this thesis we will therefore concentrate on contrastive theme–rheme pairs, which exhibit a contrast in both theme and rheme. This can be captured with Steedman's theory, which will be our point of reference for the experimental chapters to follow.

Chapter 3

Production Experiment

In the semantic literature on the formalisation of contrast it is assumed that contrast is a binary notion that is retrievable from the surface structure. Whether this is word order or prosodic realisation, the criteria are often not discussed very explicitly (cf. Rooth, 1992; Büring, 1997)[1]. Steedman (1991, 2000), a noteworthy exception, states clearly what pitch accents and boundary tones he assumes to signal theme and rheme (following Halliday (1967) in relating information blocks to tone groups). Furthermore, accented material plays an important role in his theory since it is formalised in terms of alternative sets (as described in section 2.2.2). With respect to *thematic* material, which is of particular interest for the present thesis, he takes it for granted that intonational marking expresses contrastiveness[2]:

> "It is only appropriate to mark the theme with an L+H* pitch accent when it stands in contrast to a different established or accommodatable theme. [...] Unmarked themes do not contrast with any earlier one; that is, they are entirely given or background." (Steedman, 2000, p. 656)

In the remainder of this thesis, we will use Steedman's approach to point out the difficulties associated with an assumed 1:1 mapping from intonational realisations to contrast. While we agree with the view that contrastive themes need to be intonationally marked (at least in languages like English and German), we don't think that the reverse is necessarily true: observation of real speech data shows that pre-verbal material in German often bears a pitch accent even if there is no contrast to a different theme. Even if this might appear to partially contradict Steedman's thesis at first sight, he actually covers this by using the expression "accommodatable theme". With the term "accommodation" (Lewis,

[1] In a later publication Büring associates the contrastive topics with the B accents introduced by Jackendoff (1972), though, a fall-rise accent, which — in AM terms — would be an H* LH% or L+H* LH%, see Büring (2002).

[2] Note that Steedman equates "unmarked" with "not bearing a pitch accent". This association seems appropriate if we see "marked" as specified with a feature [+acc] (Trubetzkoy, 1939).

1979; Thomason, 1990), he refers to listeners' ability to "repair" an actually non-contrastive context when encountering an intonationally marked theme. In other words, the context is then interpreted as contrastive and the utterance hence becomes felicitous to the context. Such "loopholes" are necessary for theories of information structure because they have to account for cases of "implicit contrast" that is only salient for the speaker, but which may be retrieved by the listener. More generally, Thomason points out that "[i]f there is such a thing as accommodation, genuine pragmatic rules will be routinely violated, and these violations will not seem anomalous; they will achieve an effect that seems natural and normal. Unfortunately, this relaxation will loosen the connection between theory and evidence" (Thomason, 1990, p. 333).

However, we do not share Steedman's view that these prenuclear accents necessarily have a contrastive function and this was one of the starting points for the production experiment described here. In this chapter we shall present a data collection that allowed us to investigate the phonetic properties of German themes in contrastive and non-contrastive contexts. Unlike experiments in earlier studies (e.g. Mehlhorn, 2001), we did not concentrate on semantic textbook examples. Mehlhorn, whose main purpose was to elicit hat patterns, mainly investigated utterances with sentence-initial quantors or object noun phrases (in addition to some neutral declaratives). For such examples with a very marked (uncommon) word order, intonational emphasis on the pre-verbal constituent is essential for processing (cf. Steube, 2001). Instead, we used neutral declaratives, which have no a priori contrastiveness. This allowed us to isolate the influence of context on the intonational realisation.

In the following section, we first motivate our main hypothesis, before we describe the experimental method, the stimuli used, and the recording procedure (section 3.2). From section 3.3 onwards, we shall discuss the results of the experiment. For a better understanding, we first provide a qualitative description of the differences found in contrastive and non-contrastive contexts, discussing some examples. In section 3.4 we briefly summarise how the data were preprocessed and labelled before we proceed to the results of a statistical comparison (section 3.5). Finally, we discuss the pros and cons of the method used and finally summarise the results in section 3.6.

3.1 Hypotheses

Our main hypothesis was that semantic theories oversimplify if they assume that the binary distinction between contrastive and non-contrastive themes is solely signalled by a distinction based on [+/- acc] (as assumed by Steedman, 2000). First of all, the prenuclear accents that Steedman self-evidently assigned a contrastive function might simply be caused by rhythmic considerations (secondary accents, cf. Ladd, 1996)[3]. Moreover, there seems to be an asymmetry in articulatory precision and effort in favour of initial segments (cf. Fougeron and Keating, 1996). A phrase-initial pitch accent could thus also be interpreted in the realm of phrase-initial strengthening, which signals internal discourse structuring. This marking is necessary because continuous speech lacks typographic information to clause and sentence separation as well as topic change (semicolons, periods, and paragraphs)[4].

Féry (1993) and Büring (1997) distinguished between two different sorts of rising accents for German[5], in Büring terms "Topic accents" vs. "secondary Focus accents" (p. 52), for Féry hat pattern I and II (p. 149ff). Impressionistically, Büring (1997) describes this difference as follows:

> "In normal speech, secondary accents and Topic accents might become indistinguishable. However, the shift in pitch connected with secondary accents usually has a smaller range than that connected with Topic accents [...]. In particular, one can force the Topic accent reading by intonating a rather major pitch shift, musically speaking at least a minor third." (p. 52).

Are these differences robust enough to be used to differentiate contrastive from non-contrastive prenuclear accents? Mehlhorn (2001) investigated the production and perception of topical elements in contrastive and non-contrastive contexts and found that topics that appeared in contrastive contexts exhibited different phonetic properties from topics in non-contrastive contexts. The main differences were a greater f_0-excursion of the rising accent as well as longer durations for contrastive topics.

However, the qualitative nature of her descriptions and the lacking statistical backing does not allow further inferences to be drawn. Her report furthermore remains unspecific with respect to the set of utterances used; she only reports on neutral declaratives in two dialogue contexts as well as utterances from the linguistic literature (p. 41). It is furthermore unclear how many readers participated. Therefore, we set out to replicate and extend the experiment by Mehlhorn (2001) and to evaluate whether the differences found were statistically robust. Mehlhorn compared the realisations in contrastive and non-contrastive context by averaging the fundamental frequency contours for both conditions. In the present work, we will use methods employed in recent alignment studies, i.e. studies that aim at finding stable segmental anchor points for the low and high targets (see section 2.1.1). In addition to a larger pitch range and longer segmental durations, we expected that the trough

[3]Steedman (1991) also acknowledged the presence of multiple pitch accents, to which he assigned a contrastive rhematic function, however.

[4]A common prosodic strategy to mark new sentences is a pitch range reset at the start of new utterances, which has been successfully used for automatic sentence and topic segmentation (cf. Shriberg et al., 2000).

[5]Steedman (personal communication) has meanwhile expressed his belief that — for German — he rather assumes a distinction between L+H* for non-contrastive and L*+H for contrastive themes.

before the actual prenuclear rise would be more pronounced in contrastive contexts (Jacobs, 1997; Büring, 1997). Fox (1984) e.g. observed that a *fall-rise* was a stronger version of the rise than a simple *rise* and was used to introduce a new topic or to contrast it to a topic that has been mentioned previously (p. 75).

Although there were explicit hypotheses to test (mainly whether there was a significant difference between thematic material realised in contrastive and non-contrastive contexts), the production experiment conducted had an additional exploratory part: We were interested in the acoustic variables that best determined the difference, if any.

3.2 Method

Many studies that investigate the prosodic properties of information structure employ a question-answer pair methodology for controlling the information structure of the test utterances (cf. Andreeva and Avgustinova, 1999; D'Imperio, 1999; Frota, 2002; Oliver and Andreeva, to appear, among others). While such a setup allows for a high degree of control, the purpose of the experiment can hardly be hidden from the informants so that we might expect exaggerated realisations. Furthermore, question-answer pairs have a severe drawback for the investigation of thematic material: In order to count as given in the test utterance, thematic material has to be made explicit in the question already. In such a scenario, however, the most natural answers are elliptic or contain reduced grammatical forms (pronounds, for instance). The realisation of such answers, however, can be hardly compared with the realisation of full forms in contrastive contexts. And forcing subjects to use non-elliptic and non-reduced answers might obscure results.

We consider it less intrusive to ask subjects to read paragraphs in which the target utterances appear in either a contrastive or non-contrastive context[6]. This experimental setup — with a sufficient amount of distractor material — better masks the *purpose* of the study (which ensures more naturally produced utterances than in the predictable method of question-answer pairs) and is also better suited to elicit *thematic* material. We are aware of the unnaturalness that might arise from the use of read speech, which obviously limits the generalisability of any results thus obtained. However, we regard read speech a valuable material to form hypotheses, which can be tested in subsequent studies using more spontaneous speech data.

Given the interest of the present study, controlling the (contrastiveness of the) context was the only possible method to influence subjects' productions. However, to reliably investigate the phonetic impact caused by contrastive and non-contrastive contexts, we had to ensure that the presence of contrastive elements was easy to detect. Of course, this context-setting method could not preclude that — in non-contrastive contexts — speakers produced

[6] For a concise summary of the production study see Braun and Ladd (2003).

versions that sounded rather contrastive. There still is the possibility that speakers construct a contrast to the elements already present in their mental model (which we cannot reset before the experiment). We assumed, however, that the reading task was sufficiently demanding to diminish extra-textual influence on the productions. The opposite unwanted scenario would be that speakers are not able to detect the textually coded contrast so that themes in contrastive contexts are not sufficiently signalled.

An important question was the normalisation of speaker-dependent influences on prosodic realisation. These inter-speaker differences are not only reflected in the average pitch height and the range used, but may also surface in a different temporal organisation of pitch accents (Monaghan and Ladd, 1990; Ladd et al., 2000; Grabe, 2002; Peters et al., 2002; Kraayefeld, 1997). There are different ways to reduce this inter-speaker variability: Mehlhorn, e.g., calculated an "average intonation contour" (cf. Mehlhorn, 2001, p. 44), which factored out the differences. Since we were interested in comparing the *actual* acoustic differences caused by the different precontexts, we decided for a within-subjects design. That is, each speaker had to read both versions of a given target utterance, once embedded in a contrastive and once embedded in a non-contrastive context paragraph (resulting in "minimal pair" utterances). This within-subjects design obviously increased the demand for distracting subjects from the purpose of the experiment, so that the presence of identical sentences in pragmatically different paragraphs had to be carefully hidden in order to avoid subjects exaggerating the differences. We decided to use two different sorts of filler paragraphs, which were presented together with the target paragraphs. First, we constructed genuine filler paragraphs (see Appendix A.3) which were not related to the target paragraphs. Second, we created a special distractor paragraph for each target paragraph, which was on the same topic as the target paragraphs (see Appendix A.3). These distractor paragraphs contained distractor *utterances* with a similar propositional content but with a different surface structure as the target utterances. An example is provided at the end of this section (see page 61).

Material

We used twelve target utterances with comparable sentence-lengths and syntactic complexity. Since we were interested in the influence of the context on the prosodic realisation of the target utterances, these did not contain lexical or syntactic markers that could be interpreted as contrastive. That is, we only used neutral, unmarked word orders, such as initial subject noun phrases (subj-NPs) that served as a subject for either a full verb or for an existential sentence and prepositional phrases (PPs), such as locative and temporal adverbs. These different word orders were chosen to collect a larger variety of data[7].

[7]These word orders do not constitute a further independent variable besides contrast. A factorial design was not possible because we did not find sufficient material with the same segmental structure in different

Target words were mainly chosen because of a sonorant structure (to facilitate acoustic analysis) and because of the presence of easily accessible alternatives. An initial idea was to have target words with three different stress positions, stress on initial, second and third position (the position of the stress is calculated with respect to the *prosodic word*, i.e. including articles and prepositions, which rhythmically belong to the actual target word). It turned out, however, that the acoustic analysis (see section 3.4.3) was hardly possible for target words with word-initial stress, such as "Lämmer" (*'lambs'*), "Jungen" (*'boys'*), "Lilien" (*'lilies'*), "Norweger" (*'Norwegians'*), and "Marion" (personal name). The problems were mainly caused by sentence-initial glottalisation, which made pitch extraction often impossible. To avoid word-initial stress, some nouns were preceded by definite articles, which gave them a generic reading (*'the Romans'*, *'the Norwegians'*, *'the lambs'*, etc.[8]). Problems with pitch extraction also emerged when short stressed syllables started with a voiced obstruent (e.g. /b/). Sometimes, even the presence of an unvoiced onset consonant in the *preceding* syllable caused problems ("Simone", "Kanaren").

Because of evidence about an interaction between the scaling and positioning of subsequent pitch accents (Terken, 1991; Ladd, 1994; Prieto et al., 1995; Ladd and Schepman, 2003), the utterances were constructed as to avoid a different rheme accent position in contrastive and non-contrastive context. However, some utterances seemed to be highly unpredictable with respect to the position of the nuclear, rhematic accent (especially ones with adverbs). Cases, in which speakers realised the rheme accent on different words were not analysed.

The stressed syllable of almost all target utterances was an open syllable with a long vowel, CVV (except for "die Lämmer" and "im Januar", where the first is especially unfavourable because it ends in an ambi-syllabic coda consonant). In most cases, the post-stressed syllable had a CV-structure (except for "die Malayen" and "in Bayern"). The final stimulus-set is shown in Table 3.1, together with the grammatical relation of the thematic constituent and the position of the word stress.

The target utterances were presented in paragraphs that set a contrastive or non-contrastive context. The mean number of sentences was comparable in both conditions, 5.5 for non-contrastive and 5.6 for contrastive contexts. Results of an experiment by Sluijter and Terken (1993) showed that Dutch listeners were able to tell which part of a paragraph an utterance was selected from. Since, in that experiment, this could be conveyed by prosody only, the relative position of the target utterances in the paragraph was controlled. Because of phonetic considerations (glottalisation, initial strengthening, final lowering), target utterances did not appear in paragraph-initial or paragraph-final position.

word orders, except for "die Armenier" vs. "in Armenien". The same holds for the variation in stress position, which is discussed below.

[8] Generic noun phrases qualify as topical elements, see Kuno (1993), p. 41.

Grammatical Relation	Stimuli	position of stressed syllable
subj-NP (agens)	Marlene spielt Klavier und kann singen.	2 (penult)
	Die Lämmer haben Angst vor Menschen.	2 (penult)
	Die Malayen leben von der Landwirtschaft.	3 (penult)
	Die Norweger geben gern Parties.	2 (antepenult)
existential	Die Kanaren sind ein Wanderparadies.	3 (penult)
	Italiener sind sehr gastfreundlich.	3 (penult)
	Die Römer waren sehr organisiert.	2 (penult)
	Die Maler sind viel im Freien.	2 (penult)
	Die Mormonen sind sehr traditionell.	3 (penult)
modifier-PP	In Bayern beginnen die Ferien Ende Juli.	2 (penult)
	In Milano kann man gut einkaufen.	3 (penult)
	Im Januar ist es frostig.	2 (antepenult)
	In Armenien schreibt man lateinisch.	3 (antepenult)

Table 3.1: Target utterances of the final stimulus-set, together with information about the grammatical relation of the constituent in the pre-verbal field and the stress position of the prosodic word (i.e. inclusive of prenominal articles and prepositions).

Non-Contrastive Context. Earlier we discussed that question-answer pairs are ill-suited for investigating the realisation of non-contrastive thematic material because they strongly favour weak proforms or ellipses. Weak proforms are also favoured in texts where the topic is led through the whole paragraph. In order to make full lexical forms plausible, we decided a) to resume a topic that was introduced at the start of the paragraph and which was followed by some background elaboration or b) to use an easily inferable relation (such as Malaysia – the Malaysians). In the examples below, we present the English translation of two paragraphs, one containing a target utterance with an initial subject-NP and one starting with a prepositional phrase. The introduction of the topic as well as the target utterance is highlighted using italics (the original paragraphs are listed in Appendix A):

Subject noun phrase: 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.

Prepositional phrase: Last year in *January*, an Australian friend came to visit me. It had already snowed and the whole landscape was white. *In January it is frosty.* For my friend, who came from the Southern hemisphere, this was particularly hard because from where she came it was summer. However, she really enjoyed the snow and we spent much time outside.

Contrastive Context. In order to set a contrastive context for the target utterances, we used contrastive theme–rheme pairs (as discussed in section 2.3.2). This is in line with the criteria established by Prevost (1995, 1996): He argued that either the use of two contrasting pairs of discourse entities (of the same type) or one pair of contrasting items together with a contrasting functor (e.g. verb) was a sufficient condition for establishing contrast (see also Theune, 1999). Since we were not interested in contrastive predicates, we designed paragraphs in which a proposition about an alternative element (or location or time) was found prior to the target utterance. However, we deliberately did not use parallel (syntactic) constructions to avoid exaggerated realisations but semantic parallelism (cf. Theune, 1999; Pulman, 1997). Furthermore, the proposition about an alternative item was mainly separated by one or two clauses from the actual target utterance. In order to make the alternative set salient, the relation between the target word and its alternative was explicitly stated in the precontext. In the examples below, both the target utterances as well as the contrasting elements are highlighted using italics. In both utterances there is a sentence signalling the relation between the alternatives (*'neighbouring countries'* vs. *'winter months'*):

Subject noun phrase: *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.

Prepositional phrase: An Australian friend, who intended to come to Europe for Christmas, asked about the temperatures in the *winter months. December* is comparatively mild in Germany; snow on Christmas has become rather rare. *In January it is frosty.* This year we often had -20 degrees C at night. In February, there may already be some spring-like days.

Distractor Paragraphs. Distractor paragraphs represent one kind of filler paragraphs. The distractor paragraphs contributed most to disguising the presence of identical utterance pairs. For further illustration, below we give a translation of the distractor paragraph for the PP target paragraphs shown above. The distractor utterance is emphasised in italics. Note that the surface structure differs from the target utterances.

The coldest month of all is January. Although the days gradually become longer, temperatures only rarely rise above zero degrees. Interestingly, it is in January when many fashion shops already change their stock to the spring collection.

Recording Procedure

Twelve native German speakers participated in the recording. They were between 23 and 35 years old. At the time of the recording they were graduate and postgraduate students, or staff members of Edinburgh University who had been in Edinburgh for periods ranging from a few months to four years. Speakers all spoke Standard German but originated from different parts of Germany (eight northern Germans speakers and four southern German speakers, for more detail see Table A.1 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. They were not paid for participation.

Subjects were recorded in a sound-proof room in the Linguistics Laboratory of Edinburgh University. The overall recording procedure lasted between 20 and 30 minutes. Subjects were given a pile of 52 A5 cards that contained short paragraphs with five to six sentences on average. The presentation of the paragraphs was block-wise randomised, i.e. two target paragraphs were separated by two fillers and we ensured that there was sufficient material between the contrastive and non-contrastive version of a given target utterance. Subjects were given a written instruction to read the utterances at normal speed as fluently as possible. They were told that they could have breaks between paragraphs and that they could scan the paragraphs before reading them aloud. I obtained their permission to stay with them in the recording room to ask for repetitions of paragraphs in cases with too many wrong starts or slips of the tongue. Data was recorded simultaneously on DAT-tapes (sampling rate 44.1kHz) and on a PC (sampling rate 16kHz).

3.3 Qualitative Results

All themes were intonationally marked with a pitch accent, irrespective of whether they occurred in contrastive or non-contrastive contexts. Therefore, all differences between themes produced in contrastive and non-contrastive context reported here merely alter the *kind* of intonational marking. Since the study was rather exploratory, we shall first give an impressionistic qualitative description of some of the obvious differences before we discuss the results of the statistic analyses.

Speakers used various means to signal the textually encoded contrast. In this section, we shall illustrate and discuss some of the contours to make the discussion more concrete. We shall mainly focus on durational and f_0-features that discriminate utterances in contrastive and non-contrastive contexts. I deliberately use "utterances" instead of "themes" here to emphasise that contrast-marking was not always restricted to the thematic part of the utterance. Overall, it was found that especially the thematic parts of utterances in contrastive contexts were produced with a higher relative energy compared to the corresponding thematic parts in non-contrastive contexts. Since energy measures (such as RMS) are highly dependent on the distance to the microphone, these are not discussed here.

Segmental durations. Durations of the pre-verbal constituents (i.e. the themes) were perceived as being longer in contrastive contexts than in non-contrastive ones. This is not surprising, as duration is one of the suprasegmental[9] features signalling accent, besides fundamental frequency and intensity (cf. Lehiste, 1970). One such durational difference between themes in contrastive and non-contrastive contexts is exemplified in Figure 3.1. The overall utterance duration is almost identical in the two contexts, 1.65s in the non-contrastive version (upper part) as compared to 1.64s in the contrastive one (lower part). We therefore time-aligned the two versions for Figure 3.1. Interestingly, only the duration of the pre-verbal constituent was stretched. Since the overall utterance duration is no different, we may therefore conclude that the rhematic part is realised shorter. That is, the "loss of time", which is caused by the lengthened thematic constituent in contrastive contexts was

[9]We shall use the notions "suprasegmental" and "prosodic" interchangeably.

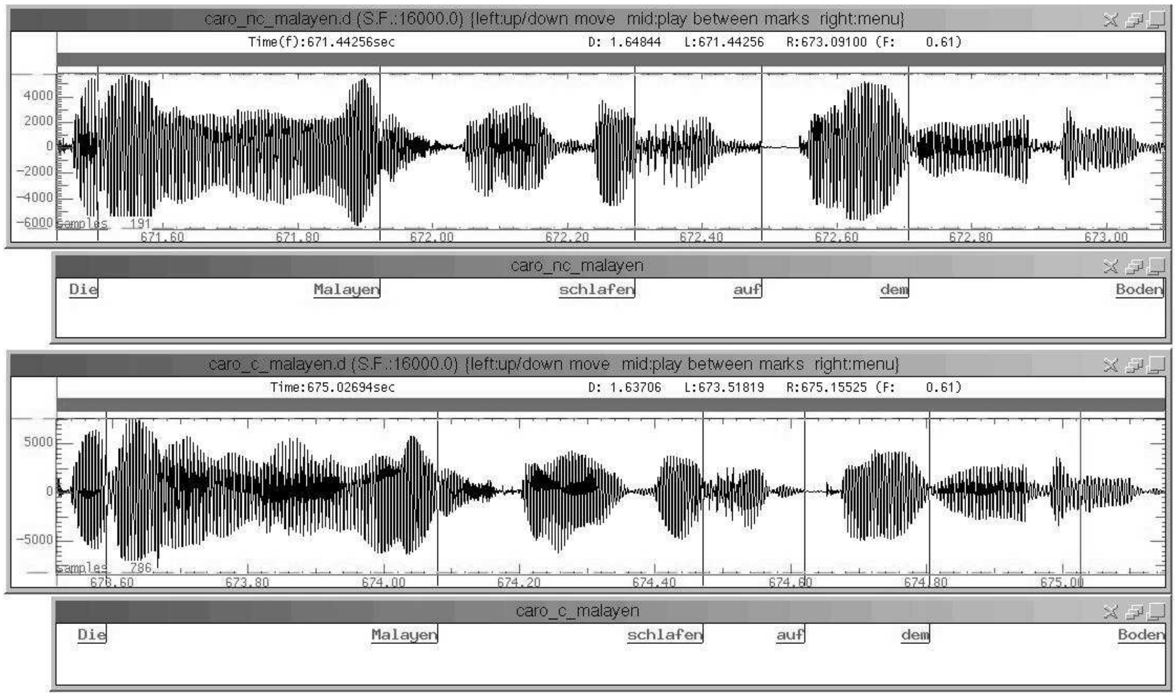


Figure 3.1: Difference in duration of the pre-verbal constituent in contrastive and non-contrastive context. Upper panel shows the utterance "Die Malayen schlafen auf dem Boden" (*'The Malaysians sleep on the ground'*) in non-contrastive context, lower panel represents the same utterance spoken in contrastive context. Word labels appear at the end of the respective words. Start and end of the utterances are time-aligned.

balanced by a shorter duration of the rhematic part of the utterance. That rhematic parts are shorter in contrastive than in non-contrastive versions is particularly noteworthy, since the contrast was established by contrasting both theme *and* rheme. Since duration is related to perceived prominence (Portele and Heuft, 1997; Streefkerk et al., 1998), this durational imbalance might shift the prominence within the utterance in favour of the thematic part.

High f_0 at start of utterance. In some cases we found a particularly high f_0-onset in contrastive contexts, see Figure 3.2, displaying the utterance "Italiener sind sehr gastfreundlich" (*'Italians are very hospitable'*). One explanation for this strategy is that a higher phrase-initial pitch increases the perception of the fall, which, in turn, might emphasise the perceived excursion of the rise. F_0-onset and range have been successful features in automatic topic segmentation (cf. Shriberg et al., 2000).

Accent type of the rise. All thematic accents discussed consist of a fall to the accented syllable, followed by a steep rise with the peak well after the stressed syllable. This is characteristic for an L*+H-accent as described in Grice and Baumann (2002). L*+H is

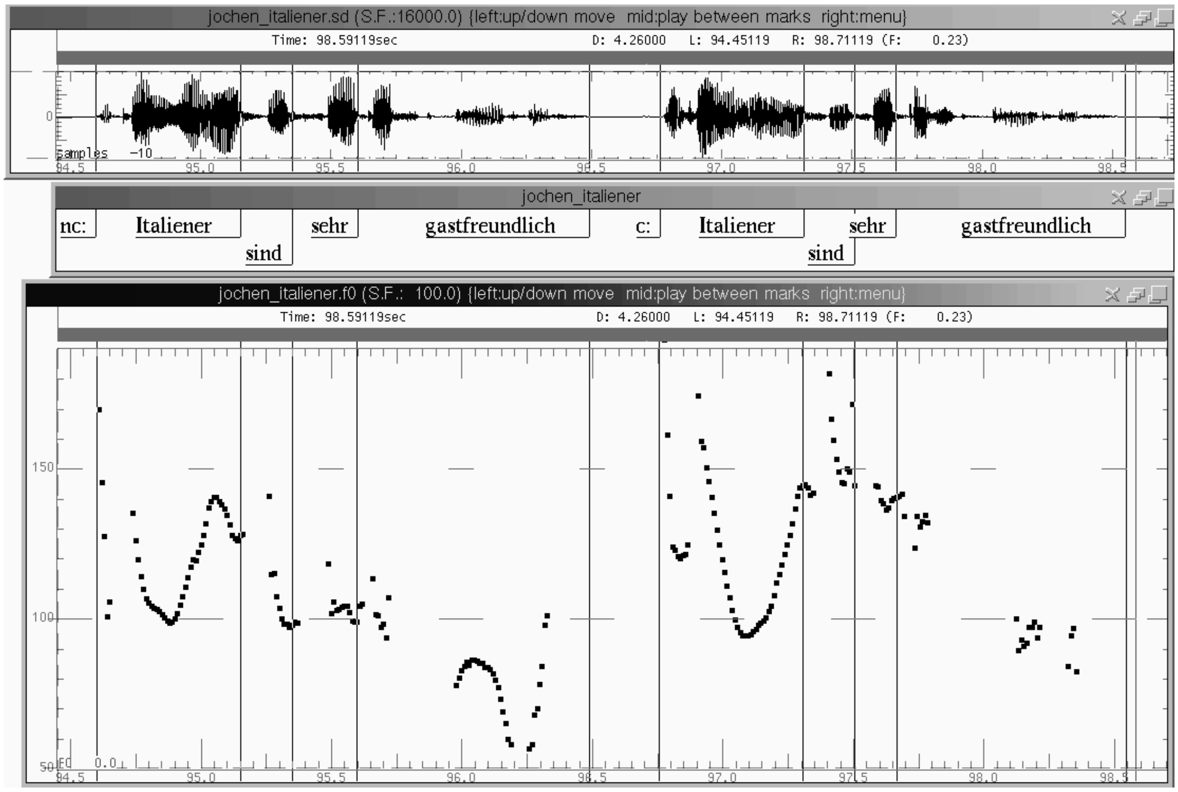


Figure 3.2: The utterance "Italiener sind sehr gastfreundlich" (*'Italians are very hospitable'*) produced in non-contrastive (right-hand side) and contrastive context (left-hand side). Note the higher f_0-onset in the contrastive context, which is difficult to detect because of the initial pitch perturbations caused by the glottal stop and the voiceless plosive.

differentiated from the other bitonal rising accent L+H* by the perceived pitch height of the stressed syllable: L+H* should be perceived as high whereas L*+H should be perceived as low. It is quite probable, however, that there is a continuum in the prosodic realisation of prenuclear accents instead of a categorical distinction[10]. Some of the accents had their minimum very early in the stressed syllable and the pitch was rising within the stressed syllable. Therefore, the stressed syllable was already perceived as rather high. It would be interesting to see whether such an accentual distinction could be related to the difference in context (contrastive vs. non-contrastive). We cannot display such examples because these fine-grained alignment differences are difficult to notice by eye. This discussion will be continued in chapter 5.

[10] The issue whether theme marking is gradual or categorical is discussed in more detail in chapter 5. Steedman (p.c.) also assumed a continuum and noted that L*+H might be a more emphasised version of an L+H*.

F_0-movement immediately after the rise and hat patterns. In most of the contours (in contrastive as well as non-contrastive context), the f_0-contour dropped after the prenuclear rise. In some utterances the contour continued its rising movement after a short dip, as illustrated in Figure 3.3 with the utterance "Italiener sind sehr gastfreundlich" (*'Italians are very hospitable'*). This continued rise might be seen as a special form of the hat pattern (first discussed by Cohen and t'Hart, 1967). A better hat pattern example is shown in the contrastive version of Figure 3.4. There the f_0 contour stays on a high level but is not further elevated as in 3.3.

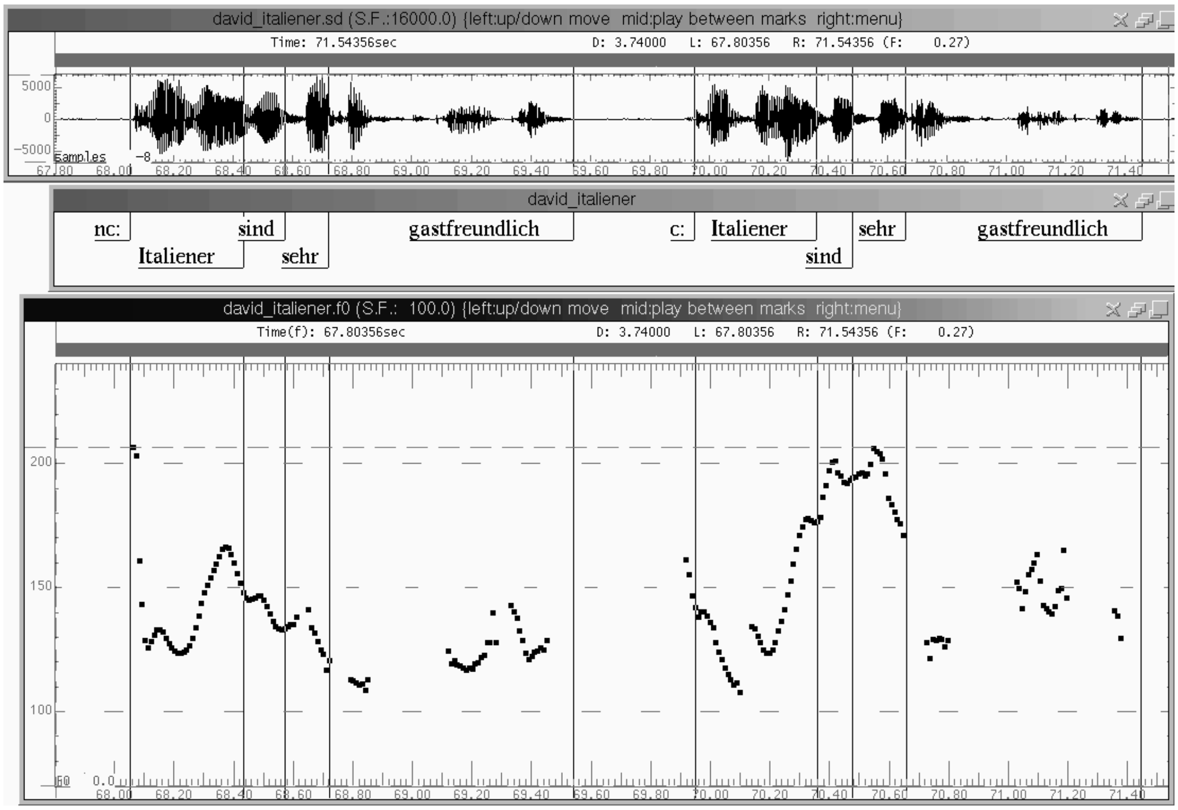


Figure 3.3: F_0-progression after the thematic rise. For the utterance "Italiener sind sehr gastfreundlich" (*'Italians are very hospitable'*) produced in non-contrastive context (left-hand side), the contour drops after the initial rise, whereas in the contrastive context (right-hand side) the rise is continued.

Inspection of pitch contours showed that hat patterns appeared in contrastive and non-contrastive contexts alike. This is in line with Cohen and t'Hart (1967) who defined this contour as the standard intonation for Dutch. Kohler discussed the hat pattern and concluded that "[t]he semantics of the most common 'hat pattern', viz. 'medial' + 'early', referring to the picking up of a theme as the basis for an observation of the obvious, makes this intonation contour ideally suited for matter-of-fact reading in German" (Kohler, 1991a, p. 328). However, hat patterns were not as frequent in contrastive contexts as could have been expected on the basis of previous literature on this topic (cf. Büring, 1997; Mehlhorn,

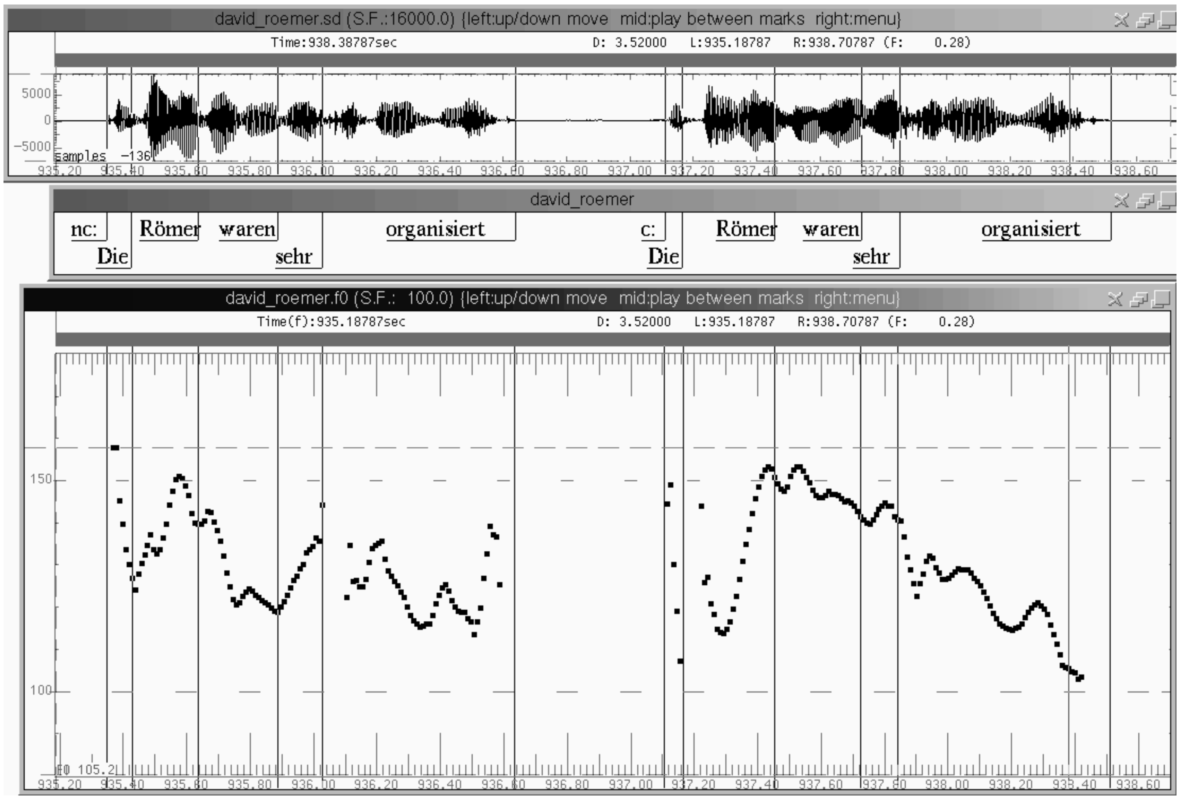


Figure 3.4: Utterance "Die Römer waren sehr organisiert" (*'The Romans were very organised'*) produced in non-contrastive (left part) and contrastive context (right part). The contrastive context triggered a hat pattern.

2001).

Rhematic accents. The nuclear accent of most utterances was realised with a (!)H+L* (or early peak in Kohler's terminology), i.e. the contour is falling to a low nuclear accent. Kohler described the semantics of this accent type as follows[11]:

> The phonetics high-low/low-high opposition signals the semantic 'closed/open' or 'established/new' dichotomy in the speaker's assessment of the communicative situation or of the speech acts. [...] Thus different F0 peak alignments with reference to the onset of a <FSTRESS> vowel express the speaker's subjective views on reality." (Kohler, 1991a, p. 306f)

According to Kohler (1991a), the predominant use of early peaks (i.e. a (!)H+L*) in our data indicates that most of the speakers judged the statements as conveying established facts rather than new information. The design of the paragraphs and the use of target

[11] In Kohler (1991a), "FSTRESS" stands for word stress.

utterances that consisted mostly of generic statements might thus have favoured these early peaks in rhematic accents.

3.4 Preprocessing and Annotation

Before analysing our data quantitatively, we had to do some preprocessing, which is described in this section. Some of the data had to be excluded from further analysis, e.g. due to hesitations (section 3.4.1). In section 3.4.2 we summarise the acoustic preprocessing, i.e. pitch correction, smoothing and interpolation, which was necessary to extract correct f_0 values. Afterwards, we give an overview over the labelling procedure and the acoustic variables extracted from the data (sections 3.4.3 and 3.4.4).

3.4.1 Data selection

The production study aimed at collecting themes that were produced in contrastive and non-contrastive contexts. While the reading task allowed us to elicit contrast "minimal pairs", i.e. identical utterances in pragmatically distinct contexts, we had to cope with the problem that not all subjects were equally good readers. Poor readers with many mispronunciations and hesitations who appeared over-employed by the reading task were excluded from further analysis. This only affected the data of two northern German speakers, SZ and SV[12]. There were speakers with average and good reading performance whose realisations, however, were not always very convincing in signalling the difference in context intonationally (speakers IB, MB), maybe because they regarded the context as sufficiently informative so that additional intonational marking was superfluous. To avoid arbitrary exclusion of data and in order to have a representative sample, we did not exclude these data.

In addition, some productions had to be excluded, because the target words were inappropriate for acoustic analysis. As already indicated, some of the ill-suited target utterances were changed during the course of the experiment. Excluded target words were those with word-initial stress (utterances "Marion", "Jungen", "Lämmer", "Lilien", and "Norwegen" of the first four speakers, CZ, CS, UB, and PK), those that were produced with additional secondary accents within the target word (four syllable word "MarmeLAde" where some subjects inserted a rhythmic secondary accent on the first syllable), utterances that provoked stress-clashes ("Die Norweger feiern gern", which led to some cases with deaccented themes), utterances where the theme was signalled with a nuclear accent type (either because of deaccentuation of the remaining utterance or by inserting ip-breaks after the pre-verbal constituent), utterances with hesitations (CZ_bayern), and utterances with a strong attitu-

[12]The data became dialectally more balanced, leaving six northern German speakers and three southern Germans.

dinal component (IB_milano, which was realised with an expression of impatience). Also, utterance pairs for which the rhematic accent was realised on different words were excluded.

As a result of this data selection, we had 86 utterance pairs for further acoustic analysis. Since we were interested in within-speaker comparisons, it was not detrimental that there were more tokens from some speakers and less from others.

3.4.2 Preprocessing

This section summarises the major preprocessing steps (and can be skipped without missing important information). First, wrongly extracted pitch values, which are artefacts of the f_0-algorithm used, were manually corrected. This mainly concerned pitch doubling and halving, and the interpolation of missing values, which occurred most often in slightly creaky female speech. Second, the f_0-contour was smoothed to reduce the influence of segmental effects on prosody (microprosody).

Data analysis was performed using xwaves, f_0-tracking with the built-in tool "get_f0", which is based on normalised cross correlation function and dynamic programming (cf. Talkin, 1995). To assure a relatively smooth pitch track, we had already used target words with mostly sonorant material. Nevertheless, we found some cases of phrase-initial pitch perturbation, which was mainly caused by glottalisation or segmental perturbation (e.g. /d/ in the definite articles "die" and "der"). Therefore, prior to analysis, a smoothing algorithm was applied to the first part of the f_0-contour. We investigated the impact of four different smoothing techniques for the use in prenuclear accents: two methods were based on median-smoothing (window size of five or seven frames, 7.5ms each) and two on mean-smoothing (again with five or seven frames). Broadly speaking, the algorithm moves a window through the speech signal and for each given frame (starting from the second or third, respectively), the actual value is substituted by the mean or median value of the whole window. Figure 3.5 shows the effect of the different smoothing techniques.

The two topmost lines in the graph, which represent median smoothing, extremely broadened the peak, so that the exact temporal alignment of the peak got lost. Comparing the smoothing with five and seven frame windows, the windows with more frames obviously resulted in a smoother pitch contour. For the analysis, therefore, a seven window mean smoothing algorithm was applied (middle contour of Figure 3.5). In order not to lose the temporal alignment of the pitch contour, the smoothed contour was reconverted into an esps-fea-file (with the command fea_edit), which allowed further analysis using xwaves.

3.4.3 Annotation

On the segmental level, four landmarks were labelled. Label points were always set at the positive f_0-crossing, using information from a wide-band spectrogram. The segmental labels

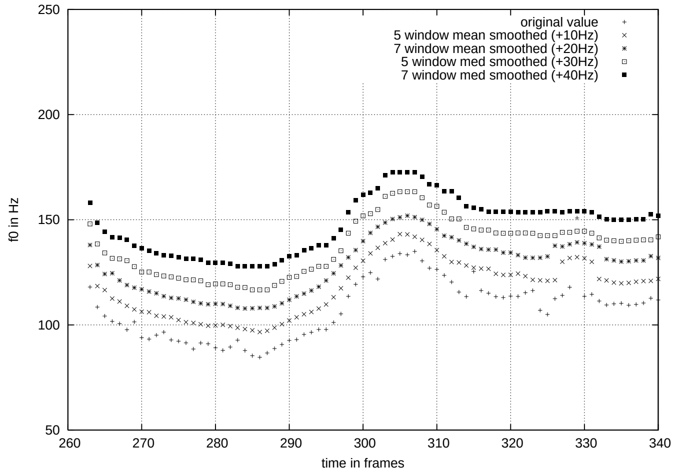


Figure 3.5: Different smoothing techniques, exemplified with a prenuclear accent of speaker JL, realised in contrastive context. Note that the order in the legend is the reversed order of the lines in the figure, i.e. the lowest line represents the original pitch contour, followed by the 5 frame mean, 7 frame mean and so on. The lines are plotted with a pitch difference of 10Hz to pull them apart.

were adapted from Ladd et al. (2000); Atterer and Ladd (2004) and concentrate on the area around the stressed and post-stressed syllable. This takes into account that the end of the prenuclear rise is often found *after* the stressed syllable:

C0: start of the stressed syllable (syllable onset)[13]

V0: start of the stressed vowel (rhyme onset)

C1: end of the stressed vowel; in most cases this corresponds to the start of the first post-stressed syllable, which was consonantal (except for “Malayen” and “Bayern”)

V1: start of the unstressed vowel following the stressed syllable

For the annotation on the suprasegmental level, we decided on a purely acoustic analysis of the pitch contour. Broadly speaking, we annotated the high point at the start of the utterance (H%), the local minimum just before the accentual rise (L), and the local f_0-maximum (H*) after the rise. Note that these labels are not meant to correspond to standard ToBI labels but are simply a notation for specifying f_0-minima and maxima. The labels are displayed in Figure 3.6. Annotation was performed by hand but could in principle be automated.

[13] Although possible according to German phonotactics, there were no complex onset consonants in our target words.

H%: High point at the beginning of the contour. In most cases this value was found in the middle of the vowel of the first unstressed syllable of the prosodic word. If the pitch values of this area were not trustworthy (e.g if there was strong phrase-initial glottalisation or devoicing), the value in the middle of the following sonorant was taken instead. This was often the case in PPs beginning with "in" or "im".

L: Local minimum before the prenuclear rise. Since the f_0-contour is influenced by the segmental structure (microprosody), spurious local minima could easily be misinterpreted as tonal targets. We did not mark cases as local minima whose f_0-values were more than 20Hz higher than a clearly identifiable trough earlier in the word. A further reason for identifying these minima as spurious was that they were all found within +/- one frame of a segmental change (example markus_roemer, jochen_italiener, but not isa_malayen). A further difficulty posed the presence of a low plateau. In order not to lose possibly important information, we additionally labelled the point just before the rise at which a considerable change in slope happened (E). In cases with a low plateau, E was marked at the last point of the plateau. This label sometimes coincided with L.

H*: First local maximum following the stressed syllable. If the peak formed a plateau, the first high point was chosen for H*[14].

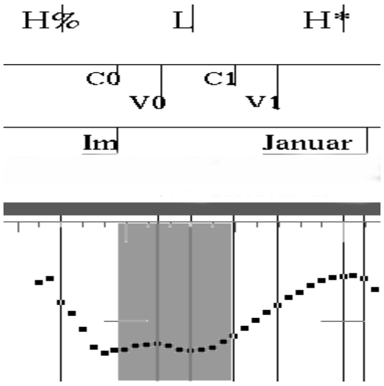


Figure 3.6: Suprasegmental and segmental labels, together with word information ("im JAnuar", with shaded stress). Suprasegmental labels mark the position of the respective event, segmental labels mark the start of the phoneme. The lexical labels, however, are placed at the end of words.

[14]There was a stronger tendency for high plateaus in words with antepenultimate stress than in words with penultimate stress but this observation will not be followed further here.

3.4.4 Description of Acoustic Variables

One aim of the production study was to explore which variables were meaningfully used to distinguish the realisation of themes produced in contrastive and non-contrastive contexts. We discuss four groups of variables, (i) pitch variables, (ii) temporal variables, (iii) alignment variables that describe the temporal association between segmental and suprasegmental structure, and (iv) the slope of the rise. Because of the difficulty associated with labelling the correct position of the local minimum, most variables that were calculated with respect to L were also calculated with respect to E:

F_0 Variables. The most obvious parameter in intonation research is f_0. We analysed f_0-variables that represent both static and dynamic properties:

f_0(H%): The absolute f_0-value before the fall to the local minimum

f_0(L): The absolute value of the local minimum (before the rise)

f_0(E): The absolute value of the elbow, i.e. at the position where the rise actually starts

f_0(H^*): The absolute value of the peak

Δf_0(fall): The range of the f_0-fall (in Hz and in semitones)

Δf_0(rise): The range of the f_0-rise (in Hz and in semitones)

Δf_0(Erise): The range of the f_0-rise starting from the elbow E (in Hz and in semitones)

Temporal Variables. We have already shown in the qualitative description of the data (section 3.3) that the segmental duration differed in the two conditions. For the statistical analysis we used two different sorts of temporal variables, ones that signal segmental durations only (from vowel duration to utterance duration) and ones that signal the duration of suprasegmental events, such as the duration of the f_0-fall and rise:

Δt(V0): Duration of the stressed vowel

Δt(syll): Duration of the stressed syllable

Δt(const): Duration of the pre-verbal constituent (= theme)

Δt(utt): Duration of the whole utterance

Δt(fall): Duration of the fall before the actual rise (from H% to L)

Δt(rise): Duration of the rise (from L to H^*)

Δt(Erise): Duration of the rise, calculated from the elbow point (from E to H^*)

Alignment between segmental and suprasegmental structure. The variables introduced above were restricted to information from one descriptive/autosegmental tier only, segmental or prosodic. The alignment variables discussed in this paragraph represent a link between the tiers, insofar as they encode the temporal alignment of suprasegmental events with respect to the segmental structure.

Recent intonation research into the association between segmental and suprasegmental structure has investigated the stability of segmental anchor points for tonal events under varying conditions. These include different speech rates, different segmental durations, as well as the distance between the respective accent and upcoming accents or prosodic boundaries (cf. Caspers and van Heuven, 1993; Prieto et al., 1995; Xu, 1998; Arvaniti et al., 1998). There has been strong agreement for the presence of *one* fixed anchor point in the segmental structure (e.g. Caspers and van Heuven, 1993; Prieto et al., 1995; Arvaniti et al., 1998; Xu, 1998). The analysis of only *one* stable anchor point (e.g. the alignment of L with respect to the start of the stressed syllable) is in line with an approach that treats the pitch accent as some sort of gesture, which takes a specific amount of time to finish (cf. Silverman and Pierrehumbert, 1990). It is furthermore compatible with a British-School approach. However, Arvaniti et al. (1998) showed for Greek that there are *two* segmental anchor points for tonal movements. This can best be explained in autosegmental-metrical terms[15], where both the start and the end of the movement can be described separately. This makes the shape (and the rising movement) an epiphenomenon. In order not to lose possibly important information, we included variables for the alignment of both the local minimum and maximum.

There is still no consensus to which segmental anchor points prosodic events are aligned. Among other indications, this becomes for instance apparent in the discussion in Ladd et al. (2000) on peak alignment in Dutch rising accents: "It may be that C1 (the end of the stressed vowel) is not the 'true' segmental anchor for the *H* and that therefore it is not the most appropriate reference point for our analysis. We therefore redid the analysis of *H* alignment, this time measuring the alignment relative to 'V1', the onset of the unstressed vowel following the stressed syllable" (p. 1550). Taking note of this problem, we calculated the alignment with respect to *different* segmental landmarks already from the outset. Since prosodic events may be found before or after a specific segmental landmark, for all variables listed below, *positive* values indicate that the prosodic event is found *after* the respective segmental position, negative values vice versa:

al(H*,V0): Duration from the start of the stressed vowel (rhyme) to the peak

al(H*,V0end): Duration from the start of the post-stressed syllable to the peak

[15] The presence of two segmental anchor points, however, undermines the theoretic claim that only the starred tone is anchored in the segmental string. More detrimental for the current definition of bitonal accents in ToBI was that the anchor points found in Arvaniti et al. (1998) lie outside the stressed syllable.

al(H^*,V1): Duration from the start of the first post-stressed vowel to the peak

al(L,C0): Duration from the start of the stressed syllable to L

al(L,V0): Duration from the start of the stressed vowel (rhyme) to L

al(L,C1): Duration from the start of the post-stressed consonant to L

al(L,V0end): Duration from the start of post-stressed syllable L

al(E,V0): Duration from the start of the stressed vowel (rhyme) to the elbow E

al(E,C1): Duration from the start of the post-stressed syllable to the elbow E

The comparison of alignment data across items is obscured if the segmental anchor points (C0, V0, C1, and V1) appear in different orders. Almost all the target words were of the structure C0–V0–C1–V1 (CVCV hereafter). There were two exceptions, however, for which the order of C1 and V1 was reversed: "Malayen" (*'Malaysian'*) and "Bayern" (*'Bavarian'*), with the structure CVVC[16]. Since we could not predict the influence of the reversed syllable structure CVVC on the alignment variables, these data were not used for the alignment parameters that describe the temporal difference of the peak to C1 and V1 (start of the onset and rhyme of the first unstressed syllable following the stress). If tonal targets are aligned with respect to the syllable structure, there should not be much difference between a syllable starting with a vowel or a consonant. Therefore, we additionally explored the alignment of the peak with respect to the start of the post-stressed syllable, irrespective of whether it is a consonant or a vowel (al(H^*, V0end)).

Derived Variables. Xu (2002) made the case for the importance of the speed of intonational changes. Therefore, we also calculated the slope of the rise:

Slope(rise): Slope of the rise, which was calculated by dividing Δf_0(rise) by Δt(rise)

Slope(Erise): Slope of the rise starting from the elbow, which was calculated by dividing Δf_0(Erise) by Δt(Erise)

3.5 Results

In the first section we shall present and discuss the results of paired t-tests, performed to establish whether the context (contrastive vs. non-contrastive) significantly influenced the variables under discussion. In section 3.5.2 we draw attention to the nature of peak height and peak alignment in our data. Finally, in section 3.5.3, we briefly discuss the use of rheme accents in the two context conditions.

[16] Note that the abbreviations CVVC and CVCV indicate the order of the segmental anchor points labelled and do not refer to the actual syllable structure.

3.5.1 Contrast *Makes* a Difference

Since all speakers read both the non-contrastive and contrastive version of each stimulus utterance, the evaluation will be strictly within-subjects. Paired t-tests were performed to investigate whether the variables differed significantly in the two contrast conditions[17]. The results for all variables are summarised in Table A.4 in the Appendix. In this section, the results are discussed in the order and grouping as introduced in the previous section. Multiple statistical tests for the relatively large number of variables derived from the same data made an adjustment of the significance level necessary. Otherwise, we would risk a Type I error, i.e. a chance occurrence of a significant difference that does not exist. The adjusted significance level is p=0.002[18].

F_0-Variables. Purely descriptively, the prenuclear rise started from a lower f_0 minimum and reached a higher peak-level for contrastive contexts. There was neither a significant difference in frequency for the height of the boundary tone (H%) nor for the minimum (L) and hence no difference in the fall before the actual rise. We do not know whether the values of these variables are genuinely overlapping or an artefact due to the difficulty in assigning frequency values to these suprasegmental landmarks. The pitch value of the elbow, $f_0(E)$ did not differ either. That is, only absolute peak height and absolute and normalised f_0-excursion of the rise (calculated from both L and from E) differed significantly between contrastive and non-contrastive contexts ($p < 0.001$). Peaks were on average 20Hz higher in contrastive contexts than in non-contrastive ones. The f_0-excursion of the rise was 1.1st (13Hz) larger in contrastive than in non-contrastive contexts.

In order to investigate which of the f_0-variables contributed most to distinguishing contrastive from non-contrastive themes, we performed a binary logistic regression with all afore mentioned f_0-variables as covariates and `contrast` as dependent variable. We used the method "forward LR selection of variables", i.e. the covariates were included stepwise and their impact on the classification was assessed. Furthermore, variables that contribute less to the classification were removed. The logistic regression showed that the f_0-excursion of the rise, Δf_0(rise)_st, was best for predicting the contrast of the precontext, with an overall correct classification rate (recall) of 65.5%. The importance of this variable is not

[17]Note that when the independent variable has only two levels, paired t-tests are equivalent to the F-statistic ($t^2 = F$).

[18]This correction of the significance level is called Bonferroni-adjustment. The standard significance level of 0.05 has to be divided by the number of independent variables tested. We have four independent tests for the absolute f_0-values of the prosodic landmarks, seven tests for the temporal variables, nine tests for alignment variables, and two for the slope of the rise, totalling in 22 tests:

$$\text{new significance level} = \frac{\text{standard significance level}}{\text{number of tests}} = \frac{0.05}{22} = 0.0023 \tag{3.1}$$

variable	#	non-contr.	contr.	p	sig?
f_0(H%)	86	168.1 Hz	167.1 Hz	0.5	
f_0(L)	86	154.9 Hz	151.9 Hz	0.01	
f_0(E)	86	157.3 Hz	155.1 Hz	0.08	
$\mathbf{f_0}$(H*)	86	218.0 Hz	227.7 Hz	< 0.001	yes
Δf_0(fall)	86	13.2 Hz	15.2 Hz	0.19	
Δf_0(fall)_st	86	1.4 st	1.8 st	0.06	
$\mathbf{\Delta f_0}$(rise)	86	63.1 Hz	75.8 Hz	< 0.001	yes
$\mathbf{\Delta f_0}$(rise)_st	86	5.4 st	6.5 st	< 0.001	yes
Δf(Erise)	86	60.7 Hz	72.6 Hz	< 0.001	yes
Δf(Erise)_st	86	5.4 st	6.5 st	< 0.001	yes

Table 3.2: Differences between f_0-variable values, extracted from themes in contrastive and non-contrastive context. Table summarises all f_0- variables, the number of pairs analysed, the respective values of these variables in contrastive and non-contrastive realisation and results of a two-tailed paired t-test. The last column indicates whether this difference is significant, based on the adjusted significance level of p=0.002 (see equation 3.1). The respective variables are printed in bold face.

surprising given that the f_0-expansion of the rise (in semitones) is a normalised variable that largely ignores speaker-dependent differences in pitch. Given a chance classification rate of 50% (contrast vs. non-contrast), the prediction of contrast does not seem rather high. Note, however, that this classification rate is achieved with not very pronounced read speech data and might be higher in natural productions.

Temporal Variables. In purely numerical terms, mean durations of all segmental variables were longer in the contrastive condition: there was a tendency for longer stressed vowels (Δt(V0)), longer stressed syllables (Δt(syll)), longer pre-verbal constituents (Δt(const)), and longer durations for the whole utterance (Δt(utt)). With our rather conservative significance level of 0.002, however, only the duration of the pre-verbal constituent differed significantly in the two contrast conditions. The thematic constituent was on average 21ms longer in contrastive than in non-contrastive contexts. Both duration of the stressed vowel and duration of the stressed syllable were close to being significantly different, see Table 3.3.

It is noteworthy that only the duration of the thematic constituent differed significantly in the two contexts. This indicates that the stressed syllable (and the stressed vowel) are lengthened roughly to the same extent as unstressed syllables. To test this hypothesis, we further calculated variables that represented the ratios of one duration variable

variable	#	non-contr.	contr.	p	sig?
Δt(V0)	86	96.4 ms	102.2 ms	0.004	?
Δt(syll)	86	152.7 ms	161.0 ms	0.003	?
Δt(const)	86	436.6 ms	457.1 ms	< 0.001	yes
Δt(utt)	86	1611.6 ms	1622.4 ms	0.32	
Δt(fall)	86	115.2 ms	125.5 ms	0.11	
Δt(rise)	86	178.8 ms	199.3 ms	< 0.001	yes
Δt(Erise)	86	156.2 ms	173.1 ms	0.001	yes

Table 3.3: Differences between the temporal variables, extracted from themes in contrastive and non-contrastive context. The table summarises all temporal variables, the number of pairs analysed, the respective values of these variables in contrastive and non-contrastive realisation, and the results of a two-tailed paired t-test. The last column indicates whether this difference is significant, based on the adjusted significance level of 0.002. The respective variables are printed in bold face. Question marks indicate that the difference was "almost significant".

with respect to the next larger duration unit. More specifically, there was one variable representing the ratio between the duration of the stressed vowel and the duration of the stressed syllable (V0:syll), one for the ratio between the stressed syllable and the pre-verbal constituent (syll:const), and finally one for the ratio between pre-verbal constituent and utterance (const:utt). Paired t-tests for these variables showed that only const:utt differed significantly in the two contrast-conditions ($p < 0.001$). That is, the duration of the pre-verbal constituent *with respect to overall utterance duration* was significantly longer in contrastive contexts than in non-contrastive ones (mean ratio of 27.2% for non-contrastive context compared to 28.3% for contrastive contexts). The domain of lengthening was the whole thematic constituent; the stressed vowel and the stressed syllable were not lengthened proportionally stronger than the unstressed syllables in the contrastive theme.

Turk and White (1999) similarly found that accentual lengthening in English showed relatively large rightward effects within a word. In other words, accentual lengthening was not restricted to the stressed syllable only but also affected following unstressed syllables. In addition they also found slight leftward effects but only for some speakers. The fact that the variable syll:const did not differ significantly deviates, however, from the findings of Turk and White (1999). This difference in results might be explained by the fact that these studies compared accented to unaccented realisations while we compared two different kinds of accented constituents.

The finding that utterance duration did not differ significantly in contrastive and non-contrastive contexts, but the derived variable const:utt did, furthermore showed that the thematic lengthening due to contrast was at the expense of rhematic duration[19].

Apart from these segmental durational patterns, there was a significant difference in the duration of the f_0-rise, which was longer in contrastive contexts than in non-contrastive ones, irrespective of whether the duration was calculated from L or the elbow point E. In contrastive contexts, it took on average 21ms longer to reach the peak than in non-contrastive contexts.

Alignment Variables and Slope. All H*-alignment variables differed significantly in the two contrast conditions, which suggests that the effect is fairly robust. The peak was consistently later in contrastive contexts than in non-contrastive ones, on average about 20ms (depending on the segmental reference points chosen)[20].

variable	#	non-contr.	contr.	p	sig?
al(H*,V0)	86	169.5 ms	195.3 ms	< 0.001	yes
al(H*,V0end)	86	73.0 ms	93.1 ms	< 0.001	yes
al(H*,V1)	70	20.9 ms	42.4 ms	< 0.001	yes
al(L,C0)	86	47.0 ms	54.9 ms	0.16	
al(L,V0)	86	-9.3 ms	-4.0 ms	0.30	
al(L,V0end)	86	-105.7 ms	-106.2 ms	0.93	
al(L,C1)	70	-105.5 ms	-103.6 ms	0.75	
al(E,V0)	86	13.3 ms	22.1 ms	0.06	
al(E,C1)	70	-81.8 ms	-75.3 ms	0.19	
slope(rise)	86	0.35	0.38	0.02	
slope(Erise)	86	0.38	0.42	0.005	?

Table 3.4: Results of the two-tailed paired t-test for the alignment values and for slope. The table summarises all variables, the number of pairs analysed, the respective values of these variables in contrastive and non-contrastive realisation, and the results of a two-tailed paired t-test. For the alignment variables, positive values indicate that the tonal event is found after the segmental reference point, negative values vice versa. The last column indicates whether this difference is significant at p=0.002, a question mark indicates an "almost significant" result. The respective variables are printed in bold face.

From Table 3.4 it becomes clear that — on average — the peak was found within the post-stressed (unstressed) syllable. Such late peaks are rather widespread, at least for

[19]This was indicated in the qualitative analysis, see section 3.3.

[20]Silverman and Pierrehumbert (1990) argued that relative alignment variables, calculated with respect to the duration of the stressed syllable, are more informative than absolute alignment variables. We hence also calculated relative alignment variables, which showed the same effect.

prenuclear accents: Silverman and Pierrehumbert (1990) found that the f_0 maximum that is associated with a prenuclear H* pitch accent in English often occurred after the syllable carrying the pitch accent. Similarly, in Korean, the tone associated with the initial accent often occurred after the stressed syllable (unless the syllable was immediately followed by another stressed syllable in the following word, cf. de Jong, 1994). Arvaniti et al. (1998) found for Greek that "the H target is very precisely aligned just after the beginning of the first postaccentual vowel" (p. 23). These studies, however, were mainly concerned with the question of peak alignment and did not discuss that peaks could be delayed with the purpose of signalling a certain linguistic function, as seems to be the case here.

There was no significant difference in the slope of the rise. If we calculate the slope of the rise from the elbow point, the difference almost becomes significant, however, with contrastive contexts having a steeper slope than non-contrastive contexts.

Again, we performed a binary logistic regression, this time to investigate which of the three peak-alignment variables was best for predicting contrast. Because of the variable al(H*,V1), we could only include the 70 cases with a CVCV structure. It turned out that the temporal difference between the peak and the start of the rhyme, al(H*,V0), was the best variable to predict the contrast (classification rate: 61.4%). We assume that the importance of al(H*,V0) can be explained by a combined effect of peak delay and segmental duration (the influence of segmental duration becomes larger, the further away the segmental anchor point is).

3.5.2 On the Nature of Peak Height and Peak Alignment

Most of the differences in thematic accent marking could be expected from the introspecitve descriptions found in the literature, except for the later peak in contrastive contexts. Gussenhoven (2002) argued that higher peaks are later because it takes a longer time to reach them. Xu (2002) and Arvaniti et al. (1998), on the other hand, argued that articulatory velocity can be increased to achieve higher targets. We were thus interested in investigating whether the delay of the peak was correlated with the height of the peak or whether these two variables were adjusted and controlled independently. First, we calculated Pearson correlations for the raw data, i.e. contrastive and non-contrastive versions were pooled. We found that peak height significantly correlated with the slope of the rise ($r = 0.82$, $p < 0.001$, $N = 172$) and with the excursion of the rise, but much less ($r = 0.40$, $p < 0.001$, $N = 172$). The height of the peak was not correlated with the position of the peak (for none of the alignment variables measured).

Obviously, the raw data are subject to much unwanted variation, such as different f_0-level or speech rate. So we calculated another set of correlation analyses, but this time the values are based on 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. For peak height ratio, the absolute peak height in each contrastive version was divided by the peak height of each corresponding non-contrastive version. Similarly, f_0-excursion ratio was calculated, which — according to the binary logistic regression — was the most important variable signalling contrast. For peak alignment,useful since some individual values we calculated the *difference* in peak alignment (positive values represent cases in which the peak was later in contrastive versions than in non-contrastive ones, and negative values vice versa).

Peak height ratio correlated significantly (but not very strongly) with the slope ratio (Pearson correlation coefficient $r = 0.53$, $p < 0.001$, $N = 86$) and with the f_0-excursion ratio ($r = 0.60$, $p < 0.001$, $N = 86$). Again, we found no correlation between peak alignment and peak height. That is, the height of the peak was controlled independently of its position. Later peaks are thus not mere phonetic concomitants of the higher peaks but can be manipulated on purpose — and probably in a linguistically meaningful way.

Speaker Differences

One observation of the labelling procedure was that some speakers made greater use of f_0-range, others of peak alignment. More specifically, there appeared to be a trade-off between the use of peak height and peak alignment, which points to some sort of "trading relation" (borrowing this term from research on perceptual cues, cf. Repp, 1982). However, we do not intend to claim yet that the perceptual impact of peak height and alignment is comparable and that a difference in encoding is not perceivable, which was found for segmental categories. We might hypothesise that the different strategies can be explained by the different dialect origins of the speakers: Atterer and Ladd (2004) reported, for instance, that Southern German speakers tend to have a later alignment than Northern German ones. Such different strategies are not readily compatible with the finding that peak height and alignment are not correlated, however.

Since impressionistic descriptions about speaker "strategies" are difficult to analyse, we first decided on objective measures that can be visualised and compared. We used the measures peak height ratio and peak alignment difference, as defined above. Figure 3.7 shows the peak alignment and peak height differences for all utterance pairs. Utterance pairs above the horizontal line had a later peak in contrastive contexts; below this line we find the cases with an earlier peak in contrastive contexts. Right to the vertical line, peak height is higher in contrastive contexts. The scatterplot shows that most of the utterance pairs differed in both peak height and peak alignment (47.8%), shown in the upper right corner.

In calculating speaker means for these variables, we saw for each speaker whether — on average for all his utterance pairs — he realised the peak later and/or higher in the

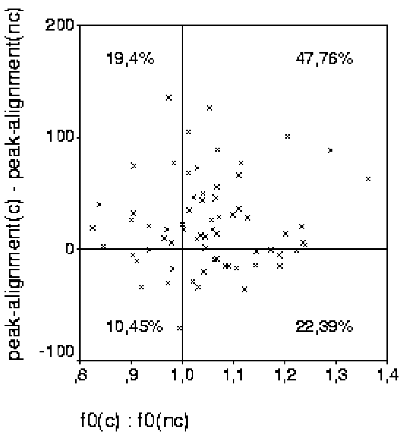


Figure 3.7: Distributional differences between peak alignment and peak height in production (all data with CVCV-structure). Difference in H-alignment in ms (calculated with respect to the start of the post-stressed unstressed vowel) between the contrastive and the non-contrastive version of each utterance pair plotted against the f_0-ratio (c:nc) of the peak. Right to the vertical line, peak is higher for contrastive versions than for non-contrastive ones, above the horizontal line, alignment is later for contrast.

contrastive versions as compared to the non-contrastive versions (and if so, how much later and higher). All speakers used a higher peak in the contrastive version; the majority of speakers used both a higher *and* later peak to signal that a theme was realised in contrastive context (see Figure 3.8). At first sight, the use of both higher and later peaks cannot be reconciled with the fact that peak height and peak alignment are uncorrelated. However, note that the individual data points seem to be negatively correlated, i.e. there are speakers that mainly used peak position for signalling contrast (especially speaker DS) and others who mainly used peak height for this purpose (see speakers BK and CS). These different speaker strategies might be the reason why — overall — some of the variables investigated (for instance the slope of the rise) did not differ significantly for contrastive and non-contrastive themes.

Although the speaker origin generally influences peak alignment (Atterer and Ladd, 2004), it does not influence the prosodic marking of contrast vs. non-contrast. Our data points, which represent *differences* in alignment and not absolute numbers, do not show an influence of speaker origin on the use of peak alignment or peak height for signalling contrast (see Figure 3.8; PK, IB, and JL are southern Germans, all others are northern Germans). However, it would be worth investigating a larger data set before drawing final conclusions.

The finding that speakers differ with respect to alignment is nothing new; it was, e.g., also observed by Ladd et al. (2000) that "speakers have slightly idiosyncratic ways of aligning"

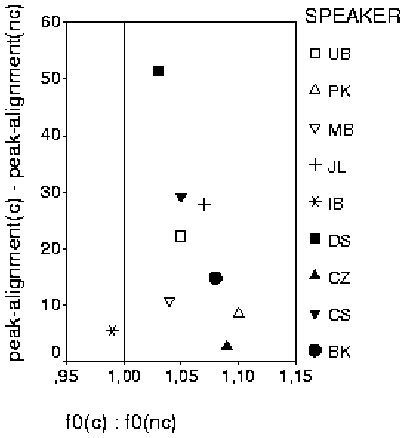


Figure 3.8: Alignment-difference in ms plotted against f_0-ratios. Data points show speaker means. All speakers used later peaks for contrastive versions as compared to non-contrastive ones. Speakers whose data points are found to the right of the vertical line used a higher peak to signal the contrastive context.

(p. 1549), which, in their experiment manifested itself as a significant main effect of speaker, which influenced the alignment of the f_0-minimum. In our case, the use of either peak height or peak alignment nicely corroborates the view that later peaks are perceptually similar to higher peaks (Gussenhoven, 2002).

3.5.3 Analysis of Rheme Accents

In the qualitative description of the data in section 3.3 we already indicated that we found utterance pairs with a different rheme accent type for contrastive and non-contrastive contexts. In this section we shall investigate whether this is statistically significant. From the literature on contrastive topics in German it was further expected that the hat pattern (i.e. an initial rise together with a sustained high pitch and a nuclear fall) was the prototypical pattern for utterances in contrastive contexts (Jacobs, 1996; Büring, 1997; Mehlhorn, 2001). Although most of our utterances were not produced with a hat pattern, we decided to investigate the frequency distribution of the nuclear accent types in more detail.

We could only perform a frequency count, because the segmental structure of the rhematic part of the test utterances was not controlled in the same way for sonority as the thematic part. That is, a similarly fine-grained acoustic analysis as done for theme accents was impossible. The use of a phonological instead of a phonetic analysis, however, is not problematic for rheme accents, since some distinctions have been shown to be categorically distinct for German (Kohler, 1991b). The categorical distinction holds between early peaks (comparable to ToBI (!)H+L*) and medial peaks (H*). Therefore, we used these accent categories, i.e. Kohler's early peaks, which are perceived as low (collapsing !H+L* and H+L*) and his medial peaks, which are perceived as high (collapsing the accent types H*, !H*, L+H*, L+!H*).

	non-contrastive context	contrastive context	total
high accent	42	26	68
low accent	44	60	104
total	86	86	172

Table 3.5: Frequency distribution of rheme accents in contrastive and non-contrastive contexts.

Overall, there were significantly more low rheme accents in our data than high ones (χ^2=7.53, df=1, $p < 0.01$). More importantly, there was a significant difference in accent type for contrastive and non-contrastive contexts (χ^2=6.22, df=1, $p = 0.013$). While there was no significant difference in accent distribution in non-contrastive contexts, in contrastive contexts there were significantly more low rheme accents than high ones (χ^2=13.44, df=1, $p < 0.01$). That is, a contrastive precontext (i.e. a contrastive theme and rheme in the precontext) provoked significantly more low nuclear accents. The overwhelming use of low rhematic accents might be responsible for the — as yet — unexplained duration distribution in the utterances found (see section 3.5.1), i.e. longer durations for thematic elements compared to shorter durations for rhematic elements although both themes and rhemes were contrasted. The use of low rhematic accents might be a more important cue for rhematic contrast than duration. Further, there might be an interaction between thematic and rhematic accent realisation, possibly a weak form of a hat pattern.

We therefore 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. Clearly, this kind of analysis is very crude, as the alignment variables are dependent on the speaking style and the prosodic structure of the word. 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 stressed syllable. We 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). Calculating a Mann-Whithney test with syllable duration (sylldur) and duration of the stressed vowel (durV0) as dependent variables confirmed our hypothesis. The durations of the thematic vowel and syllable were significantly longer when the rheme accent was low ($p < 0.01$).

It is difficult to establish a direct connection between low rheme accents and contrastive contexts as they were also quite common in non-contrastive contexts. We sympathise with the interpretation of Kohler (1991b), who described early accents (low accents in our terminology) as conveying "established facts" rather than new information (see also Ladd, 1996). The majority of early accents might thus be a mere artefact of the reading task and they might be much rarer in spontaneous speech.

3.6 Discussion

In this last section of the production chapter we will first discuss the pros and cons of the method chosen (section 3.6.1). In addition, we briefly raise an issue that has been neglected in our discussion so far, the accentuation of given information. Finally, in section 3.6.2, we summarise the main results and motivate perception studies to investigate whether the differences found are meaningful.

3.6.1 Critical Evaluation of the Method

None of the subjects reported that they had encountered identical utterances in different paragraphs. The use of distractor paragraphs on the same topic as the actual stimulus-paragraphs, together with a sufficient number of genuine filler paragraphs obviously served the purpose of disguising the goal of the experiment. Therefore, we regard our data as highly natural realisations[21], which — given the purpose of the experiment — are superior to the exaggerated versions often analysed and discussed in linguistic research.

For defining contrastive precontexts, we used contrastive theme–rheme pairs, which is in line with the criteria established by Prevost (1995), Theune (1999), and Pulman (1997). We used a proposition about an alternative element prior to the target utterance. Because of the intended comparison to identical utterances in non-contrastive contexts, contrast could not be signalled by contrastive discourse markers, so we had to find other means to ensure coherent paragraphs. We therefore often introduced the contrastive relation prior to the target paragraph. For instance, the contrast between *'Indonesians'* and *'Malaysians'* was introduced by the sentence *'Despite their geographical adjacency, their living and working conditions differ tremendously'*. Contrast hence fulfils the similarity plus dissimilarity condition formulated by Umbach (2004). In the literature, we found no operational definition of what counts as an *alternative* element. Alternatives might be defined via general knowledge

[21] Still, we are far away from real spontaneous speech, but this has to await future research.

(which could be represented in some sort of semantic ontology) or dependent on the context (for which pragmatic information is essential). Furthermore, there is no research about the impact of the *number* of salient alternatives or their respective accessibility or *salience* on prosody. Due to lacking operational criteria for alternative sets, we made possible alternatives, which we thought might be difficult to access, salient by explicitly relating them to the target already in the precontext.

While contrastive contexts can be defined in terms of a "detached" semantic parallelism, non-contrastive contexts lack such an operational criterion. Simple continuation was not possible because this would have favoured pronominalisation of the thematic constituent. So we decided to use topic resumption or easily inferable relations (e.g. *'Malaysia'*, *'the Malaysians'*). The use of discourse notions such as "topic resumption" shows that the strict binary distinction between contrast and non-contrast is too coarse-grained to describe the relation between utterances and hence different kinds of theme marking. Also, contrast can easily be merged or confounded with the discourse relations "sequence" or "enumeration" (e.g. Indonesians and Malaysians or December, January, and February).

The participants did not have a totally convincing reading performance; among other things, it was probably impaired by the unnatural situation of being recorded while reading aloud. Some subjects reported that they experienced difficulty because of the text sort (short narratives). One reader reported that he had been irritated by topic changes and that he would have preferred newspaper texts to the narratives. Speaker CZ reported after the recording that he had forced himself not to scan the paragraphs in advance to be maximally "spontaneous", which unfortunately led to some improper realisations of contrastive themes.

I silently listened to all the recordings and only asked for repetition if there were mispronunciations directly before or in the target utterance. Fortunately, mispronunciations in the target utterances were rare. Repetitions of the paragraphs could mostly not be used for analysis, because subjects tended to hyperarticulate the words that they had mispronounced beforehand, thus changing the prosodic structure of the whole utterance. For some subjects (especially speakers IB and SV) the request to repeat certain paragraphs that contained mispronunciations resulted in impatient, annoyed repetition[22].

Excursus: Intonational Marking of Given Information

For our main analysis we only examined thematic, i.e. "given" (Halliday, 1967; Chafe, 1974) material. Interestingly, but not unexpectedly, *all* themes were realised with a prenuclear accent. Comparing this to the phonetic literature on accentuation and givenness, there is large agreement that "new" information almost always has to be accented because missing accents result in perceptual inappropriateness (cf. Nooteboom and Kruyt, 1987). However,

[22]As speaker PK pointed out to me, it would have been better to ask them to repeat the respective paragraphs at the end of the recording session.

the picture is somewhat less clear for specifying the intonational marking of *given* information. Nooteboom and Kruyt conclude that "plus focus [i.e. accented, BB] can often be acceptably associated with given information — often, but not always. Again, constituent structure or word order seems to strongly influence what is and what is not possible" (p. 1520).

There have been various attempts to investigate the production of given and new information as well as the perceptual effects of accented given information (cf. Nooteboom and Kruyt, 1987; Terken, 1985, 1991, among others). Terken (1985), who worked on production, concluded that "speakers are not easily inclined to treat information as Given: even when conditions are favourable (that is, when the current item is the only one mentioned in the previous utterance, when it is salient, etcetera), speakers often use an accented expression" (p. 25). Toepel and Alter (2003) investigated the perception of contrastive and newness accents in the respective appropriate and inappropriate contexts by means of event-related brain potentials. They concluded that "[p]rosodically underspecified (focus) accents in discourse generally lead to online and offline processing difficulties, whereas overspecified ones only do so under certain task requirements" (p. 622).

These studies all investigated the intonational marking or appropriateness of *rhematic* accents, which led us to conclude that the intonational marking of non-contrastive themes was not simply a peculiarity of prenuclear accents. Probably it is easier to assume that it is not the mere distinction of given vs. new that guides the presence or absence of accents but a more complex interaction between grammatical relation, various discourse functions, etc. (cf. Terken and Hirschberg, 1994). Also, accentual marking could be guided by the difficulty of activating it. While a non-contrastive topic should be rather easy to activate, a contrastive topic has to compete against an alternative topic element.

3.6.2 Summary

The reading task was an appropriate method to elicit highly natural realisations of themes that occurred in contrastive and non-contrastive contexts. The results of the paired t-tests showed that there were indeed significant differences between non-contrastive and contrastive versions. Most importantly, themes in contrastive context exhibited higher peaks, accompanied by a greater f_0-excursion than in non-contrastive context. Furthermore, the peak was reached later in contrastive contexts than in non-contrastive ones, and the segmental duration of the target word was significantly longer in contrastive contexts. We could not verify that themes in contrastive contexts are realised with a more pronounced (deeper) trough before the actual prenuclear rise (as stated by some researchers on an impressionistic basis, e.g. Jacobs, 1997).

In addition we found that peak height and peak alignment were uncorrelated, i.e. they can be adjusted independently. This fact was clearly exploited by our speakers, who used different strategies to encode the textual contrast: They varied as to whether they used peak height, peak alignment, or both for signalling contrast. In contrast to the finding of Atterer and Ladd (2004), the choice for one of these strategies did not seem to be influenced by the speaker's origin.

The differences found in production raised the question whether these differences had an impact on the perception and hence the interpretation of thematic material in general. Especially the observed speaker strategies suggested that there is no simple intonational grammar for contrastive theme accents. Instead the analysis needs to be more fine-grained and also has to account for the speaker differences observed. A further question was whether the two factors peak height and peak alignment were equally important or whether one "cue" was stronger, more important or simply more accessible than the other one. Both questions will be addressed in chapter 4.

Another issue we haven't discussed so far is the question whether the observed differences are continuously variable or represent a categorical distinction. This issue is of great importance for semantic theories that — for their formalisation of thematic contrast — rely on categorical accent distinctions (Steedman, 2000; Büring, 2002). A mere continuous variation within the same accent type would obviously not be very supportive. Nevertheless, this view has been put forward by Gundel and Fretheim (2001). Alternatively, we might assume that thematic material is marked with a gradiently (or stepwise categorically) increasing emphasis correlated to different discourse relations (such as topic continuation, topic resumption, contrastive topics). Whether the observed differences can be reliably annotated with different phonological accent types is finally discussed in chapter 5.

Chapter 4

Perception Experiments

The results of the production experiment have shown clear differences between themes produced in contrastive and non-contrastive contexts. Furthermore, we found that subjects varied as to whether they used peak alignment or peak height or both for signalling thematic contrast. However interesting the results of such acoustic studies are, they do not reveal whether the observed acoustic differences are actually perceived by listeners[1]. This is an important precondition to regarding them as meaningful phonological differences instead of merely the result of gradual phonetic variation. This primacy of perception research has been especially emphasised at the Institute for Perception Research (IPO):

> "It is not possible, solely on the basis of a visual inspection of an F0 curve, to decide how it is converted by the auditory system into a melodic continuum, let alone that it would be possible to know what is communicatively relevant and what is not. This implies that our understanding of the listener's interpretation can only be achieved in a study dedicated to the perception of speech pitch [...]" (t'Hart et al., 1990, 24f)

In this quotation, t'Hart et al. (1990) point out the two aspects of speech pitch that are interesting for perceptual studies, "how it is converted" and "what is communicatively relevant".

In this chapter, we will first motivate and discuss the experimental setup for assessing subjects' ability to perceive the encoded contrast (section 4.1). Perception studies aim at assessing the subjective (linguistic) impressions of listeners by means of special tasks. Finding the right task is by no means trivial; it is especially difficult in intonation research because intonation can be used for many different functions which can hardly be fully controlled. We then describe the actual experiments. By means of preference tasks with resynthesised stimuli we explored the perceptual contribution of peak height and peak alignment (section 4.2)

[1]A summary of the perception experiments is found in Braun (2004).

as well as the respective influence of peak height and segmental duration (section 4.3). The results of these experiments and their possible implications are presented in section 4.5.

4.1 The Choice of Perception Method

Our main goal was to investigate whether the differences found in production had an impact on perception. Our interest goes beyond the question whether subjects are able to detect the acoustic differences: rather, we are interested in whether a given utterance can be related to a contrastive or non-contrastive context. Therefore, methods employed in similar studies meta-linguistic prominence judgements as undertaken by Mehlhorn (2001) are not useful for our purpose. There are various methods, using different sorts of stimuli, which can be employed to address this question. We shall take some time to motivate our choice for preference tasks with resynthesised stimuli. Then we summarise the foundations and similarities of the subsequent experiments.

4.1.1 Material from Production Data

A straightforward approach to answer the question whether subjects perceive the differences found in the production study, is to make use of the production material itself. One possibility would be to obtain preference ratings from cross-spliced material. In cross-splicing, the target utterance is presented both in the original and the opposite context, resulting in two matching and two mismatching conditions. This technique is rather easy to carry out in our case because whole utterances (i.e. intonational phrases) could be excised from the context, minimising segmental phonetic perturbations.

However, there are some concerns wrt to using our production data for perception experiments. First, as already discussed, not all utterances were produced as intended, which resulted in a decreased realisational contrast between themes in contrastive and non-contrastive contexts. This was caused by an insufficient interpretation of the preceding context, or by the contexts themselves not being optimally designed to make the contrast immediately obvious. In addition, there were some careful readers among our subjects who accented almost every content word and therefore made the utterances (and especially the thematic elements) clearer than necessary. Using our natural data would have forced us to preselect the stimuli, i.e. choosing those utterance pairs that exhibited clearly distinguishable realisations when judged in isolation. Although this data selection could have been performed inter-subjectively, there was a persisting danger of biasing the results. The second reason — which is closely related to the first one — is that we could not have used all of our precontexts because they sometimes contained slips of the tongue, false-starts etc. We were not confident in using these erroneous data in a rating task because we could not preclude that this had no effect on subjects' ratings. Slips of the tongue and false starts were also found in the precontexts of those utterance pairs that were judged as being clearly distinct. Therefore, excluding these stimuli — which would have been the safe way — would have reduced the amount of data by a considerable extent. Third, as the analysis of the production data has shown, speakers used different strategies to mark the encoded contrast. With a data selection as just described, it would have been difficult to ensure an equal number of versions that differ in peak height and peak alignment only. This limitation makes is almost impossible to investigate the relative perceptual importance of differences in peak height and peak position. In rating tasks with different voices, subjects might also be influenced by subjective criteria about the speaker, such as different voice characteristics (breathy or creaky), dialectal background, or articulation rate. A further important argument against cross-splicing lies in the method itself and not in our specific material: With this method, it is always difficult to be certain that positive results (i.e. good ratings for matching conditions and bad ratings for mismatching ones) result from the influence of context rather than the differences introduced by cross-splicing (slightly different register, different voice quality, speech rate etc.). The only way to reduce the influence of such phonetic effects is to reduce the overall quality of speech sounds, e.g. by LPC resynthesis.

Nevertheless, we evaluated the applicability of cross-splicing to our data by conducting two informal pilot tests, using the natural stimuli without LPC-resynthesis. First, we selected a small set of paragraphs that contained no reading errors and whose target sentences sounded sufficiently distinct when heard in isolation. to reduce the memory load, we presented the matching and non-matching paragraphs both auditorily and in written form on a computer screen. Our listeners judged all the conditions as sounding equally appropriate. When listening to the paragraphs ourselves, it was often not possible to identify the mismatching conditions — even with prior knowledge. Therefore we concluded that the simultaneous presentation of the long precontexts together with the target utterances overloads the listeners' processing abilities[2].

The unexpected result of this first pilot test raised the question why two utterances that sound clearly distinct in isolation do not affect the appropriateness of the whole paragraph. This led us to the idea of presenting both of the target utterances and asking subjects to decide which of them was more appropriate in a given precontext. The precontext in the second pilot test was again presented visually and auditorily. The results were comparable to the first pilot test; listeners appeared to assign the target sentences to the precontexts randomly. We therefore concluded that the precontext was not distinctive enough to favour one of the two realisations. This indistinctiveness of the precontext is not really surprising. In the production experiment, we intended to mask its purpose, and therefore the presence of a contrast was not explicitly highlighted. As already discussed,

[2]Another explanation is that a wide range of realisations is acceptable to listeners, i.e. that they are able to accommodate a slightly different context or speaker attitude. This issue will be brought up again in section 4.4.

we were successful in distracting our subjects from discovering the identical utterances in different contexts, but this seems to make our contexts less appropriate for perception experiments of this sort. In other words, we have to define our contexts more clearly so that subjects have a chance to build up different internal representations against which they can evaluate the appropriateness of given stimulus utterances. A different sort of perception experiments are identification tasks, in which subjects have to relate a given stimulus utterance to different semantic categories.

4.1.2 Identification Tasks

The previous discussion about the results of the informal pilot tests has shown that the contexts we used for the production experiment cannot be easily transferred to perception experiments. Before we turn to the experiments we actually performed, we would like to motivate why we abstained from using identification tasks as well. Identification tasks do completely without a precontext but instead rely on subjects' ability to *classify* the given stimuli. More specifically, there are two different variants: either each stimulus is assigned a semantic category or it is assigned a value on a scale between two opposing semantic "endpoints" (cf. Osgood et al., 1957).

In both cases, obviously, identification requires the specification of one or more meaningful categories that can be used to describe the stimulus. In the identification tasks used in classical categorical perception studies (Liberman et al., 1957), the labels were simply described as /p/ or /b/, which was sufficient. It is more difficult, however, to assign labels in the area of intonational meaning. Ladd and Morton (1997), who investigated whether the perception of intonational emphasis was categorical or not, used for example the categories "everyday occurrence" and "unusual experience" (p. 325). The success of identification tasks crucially depends on the quality of its category labelling. t'Hart et al. (1990), who doubt that intonation features have an intrinsic meaning, even argue that "the same label may have partially different meanings for each of the subjects as well as for the experimenter" (p. 112).

Finding good lables is particularly difficult for the phenomenon under study, as our discussion of the notion of contrast in the introductory chapter showed. As discussed in the introductory chapter, it is rather difficult to define the notion of contrast. Contrast is a genuinely extra-linguistic notion that relies on the conversational and situational context. A commonly used and easily describable denotation for contrast are corrections (which we argued are quite different from contrast, see section 2.3.2). For our data, however, the segmental and suprasegmental properties do not qualify the stimuli for a corrective context. For corrections, there is too little hyperarticulation on the segmental level and there is too little f_0-movement and the wrong accent type on the intonational level. Corrections are furthermore generally realised by nuclear accents, not by the prenuclear ones that we are interested in.

Semantic theories describe contrast using the notion of alternative-sets. However, this definition can hardly be used to define the notion of contrast for subjects in an identification task. We cannot imagine asking subjects whether they perceive some sort of highlighting of an alternative item to the preverbal element of the utterance. We decided, therefore, that identification tasks are not viable as a method to elicit listeners' perception of [+/-contrast]. After this methodological detour, which we consider important in justifying our approach and defining our position, we now discuss the experiments we actually performed.

4.1.3 Rating Tasks with Resynthesised Material

So far, we were mainly concerned with the benefits and shortcomings of using the natural precontexts and the method of identification tasks for our purposes. We also regarded the natural stimuli as optimal material for answering the question whether the encoded contrast is actually perceivable. However, we became more and more convinced that resynthesised material had even more advantages. These not only allow us to test whether the differences found in production are perceivable but they also help to explore the relative importance of the different acoustic variables, such as peak height and peak alignment. With resynthesised data, we can exercise a good control over the prosodic realisation while maintaining a reasonable naturalness. Furthermore, they exclude the possibility that ratings are based on subjective judgements about the assumed personality of the respective speakers[3]. We decided to use PSOLA-based resynthesis (as e.g. implemented in Praat, see Boersma, 1992).

In the following sections we shall investigate whether the differences found in the production experiment have an effect on perception as well. More specifically, we are interested in the perceptual importance of peak height, peak alignment, and the segmental duration of the preverbal constituent. In contrast to recent studies that investigate the categorical nature of certain phenomena (cf. Kohler, 1991b; Ladd and Morton, 1997; Remijsen and van Heuven, 1999; Chen, 2003), we did not resynthesise a pseudo-continuum with equally spaced steps, but simply investigated the "endpoints" established in the production data. This "endpoint-approach" is justified on two counts: First, we do not have evidence *at all* yet that there is a perceptual distinction between themes produced in contrastive and non-contrastive contexts, which would be a necessary precondition for investigating varying points between the "extremes". Second, our main interest is the *interpretation* of intonationally coded contrast for which these studies are to lay the basic perceptual foundation, namely that the intonational differences are *perceivable*. The experiments here are mainly rating studies, in which subjects are asked to judge the appropriateness of resynthesised stimuli in a given context. In this section we summarise their essential features and sim-

[3]There is some research in accommodation theory which suggests that speakers perceive and judge speech styles that are more similar to their own more favourably (cf. Street, 1982).

ilarities. In describing and motivating as much as possible in this introductory section we hope to increase the understandability of the argument expounded in the remainder of this chapter, which is finally concluded with an overview of the experiments and results (section 4.5).

Stimuli. One female speaker (not the author) read the utterances shown in Table 4.1, which are close to the stimuli used in the production experiment. They were recorded in isolation and embedded in a parallel construction (e.g. *'Anna loves the sea but Marlene likes to go to the mountains.'*), both with natural intonation and monotonously. Since the monotonous recording resulted in a rather unnatural resynthesis (especially at the end of utterances where the voice quality was not creaky enough and final lengthening was almost absent) we decided to use one of the naturally spoken utterances as a basis for resynthesis. Similar to the production study, the material we used was intentionally heterogeneous: There are target words with stress in initial, second or third position with two different grammatical relations in the preverbal field, subject noun phrases and locative prepositional phrases.

	Original stimuli	English translation
Training	Lilien sind sehr anspruchslos.	*'Lilies are rather undemanding.'*
	In Indien konzentriert man sich auf Technologie.	*'In India, technology is very important.'*
	Lämmer haben Angst vor Menschen.	*'Lambs are afraid of humans.'*
	Maurer arbeiten viel im Freien.	*'Bricklayers often work outdoors.'*
Target	In Armenien schreibt man lateinisch.	*'In Armenia, the Latin alphabet is used.'*
	Marlene fährt gern in die Berge.	*'Marlene likes to go to the mountains.'*
	In Milano kann man gut einkaufen.	*'In Milano, it is easy to go shopping.'*
	Die Norweger geben gern Parties.	*'Norwegians love to have parties.'*
	Die Römer waren sehr organisiert.	*'The Romans were very organised.'*

Table 4.1: Training and target utterances of the rating experiments

The resynthesised versions used are not exact copies of the mean values found in production, though. That is, the resynthesised rises are *idealisations* based on the production data. The deviance from the actual values was caused mainly by the higher fundamental frequency of our speaker compared to the speaker-means of the production data. If we had aimed at an exact intonational copy-synthesis (with stronger f_0-manipulation), the resynthesis quality would have deteriorated noticeably. This persuaded us to allow certain deviations from the actual f_0-values found in production.

Furthermore, the average values of the production study are based on a range of realisations, some of which were not as distinct as they could be. We dare to claim that the analysed utterances are highly natural realisations (except for the fact that we are dealing with read speech) because speakers were not aware that they read identical sentences in pragmatically distinct contexts. We therefore suggest that the production data are realisations on a continuum where the "endpoints" (if any) were only approached in some utterances. Nevertheless, there were significant differences between themes that were produced in contrastive and non-contrastive contexts. That is, although we did not analyse the extremes, the influence of context on the prosodic realisation was robust enough to surface. While this strengthened the pragmatic importance of the observed differences in production, it is unfavourable to use "in-between" realisations for perception data. There, subjects' ability to perceive a difference and to make use of it is crucial. We therefore slightly exaggerated the overall trend found in the production data, both for presumed non-contrastive and contrastive versions[4].

The experimental setup would have easily become too large if we had wanted to resynthesise all possible factors — and possibly also all interesting interactions — at the same time. To keep the experiments concise and clearly arranged, we first concentrated on the manipulation of f_0-movements, i.e. peak height and peak alignment (section 4.2). This neglected the influence of other prosodic factors such as duration, energy, and more importantly, spectral tilt that also contribute to the perception of accentuation (cf. Lehiste, 1970; Sluijter, 1995). Disregarding these other prosodic phenomena and focusing solely on f_0-movement is common practise, as an early quotation by Cohen and t'Hart (1967) shows[5]:

> "With most investigators in the field we share the common belief that reduction of the problem of intonation to a study of pitch phenomena would provide the most fruitful approach."(Cohen and t'Hart, 1967, p. 177)

Except for some few studies (cf. Heldner and Strangert, 2001; Heldner, 2003) the last 40 years haven't brought much change in this respect. Although this is a strong simplification, we initially follow Cohen and t'Hart (1967) and many others, and concentrate on the influence of f_0-movement on perception in the experiments described in section 4.2. The stimuli used in the experiments were refined step-by-step, based on the results of preceding perception experiments and also on a more focused re-analysis of the production data. That is, throughout the series of perception experiments, more and more dimensions were

[4]Note that the overall pattern of the production data was only slightly exaggerated which did not lead to easily distinguishable utterances, i.e. the differences are still rather subtle.

[5]This separation of f_0-movement from other prosodic events has entered terminology. Often, "intonation" is used for the description of changes in fundamental frequency only, whereas "prosody" covers all prosodic events, including, duration, energy and spectral tilt. This narrow interpretation of "intonation" is not used throughout the field, though.

explored. This concerns the interaction of thematic and rhematic accents (4.3.1), as well as the interaction of peak height and segmental duration (4.3.1 and 4.3.2). We assume that there is enough "overlap" between consecutive experiments — especially because changes in stimuli and contexts are well-founded — to keep the results of the experiments comparable. For an overview over all experiments see section 4.5, especially Figure 4.6.

Method. All subjects who participated in the experiments were native speakers of German (from various dialectal backgrounds) with normal hearing. They were all recruited on the university campus. Most of them were students or academic and non-academic staff, aged from 20 to 35 years. They were naïve with respect to the purposes of the respective experiments. Each subject participated in only one of the experiments described.

In all perception experiments, we presented five training utterances in fixed order to familiarise subjects with the speech stimuli and the task[6]. The target utterances were presented in random order. All the rating studies discussed in this chapter are designed as repeated measures, i.e. each subject had to judge all versions. They were explicitly told that all the versions were different and that the repeated presentation of the same utterances was not meant to test the consistency of their judgements. This design has the drawback that there might be a learning effect that cannot be controlled. However, we assumed that the randomised presentation of stimuli across subjects should be a reasonable and sufficient way of minimising such a learning effect. The other, positive side of the coin is that repeated measures are particularly advantageous for perception experiments because this method allows the inspection of intra-speaker consistency and preference patterns, at least to some extent (as will be shown in section 4.2.2, these may differ significantly in some cases). Furthermore, this design reduces the number of stimuli and subjects needed, which helped to satisfy both phonetic demands (exclusively sonorant material to enhance the quality of resynthesis) and semantic demands (noun phrases and prepositional phrases, for which alternatives can be established easily).

4.2 Influence of Peak Height and Peak Alignment

As reported in 3.5.1, one interesting difference between themes in contrastive and non-contrastive contexts was that intonational peaks were produced higher and later in contrastive contexts. Surprisingly, there seemed to be a trade-off between the use of peak height and peak position in signalling contrast.

In the perception studies described here we explored whether these acoustic differences have an influence on perception. Furthermore we addressed the question whether peak

[6]From the four available training utterances (see Table 4.1), the "Maurer"-example was presented twice to familiarise subjects with the rating of segmentally identical utterances with different pitch contours.

height and peak alignment are additive factors or whether there is a trading relation (cf. Repp, 1982). Both alternatives are possible from the results of the production data: On the one hand, in the vast majority of utterance pairs, there was both a higher and later peak in contrastive contexts, which allows the conclusion that these two factors are additive. On the other hand, we reported different speaker strategies, i.e. speakers who used peak alignment differences more extensively and others who made greater use of pitch height differences for signalling contrast. This rather indicates a trading relation.

Our experiments have only an exploratory nature because there is a strong interaction between the phonetic variables — peak height, slope of the rise, position and height of f_0-minimum, segmental duration etc. — whose causal relations are not yet clear. Resynthesising a late peak (which appeared mostly in contrastive contexts) results in a flatter rise if we do not simultaneously vary the position or the height of the trough (f_0-minimum, L). However, neither a change in the position of the trough nor in its height is directly justifiable on the basis of the production data because neither of them differed significantly. Flatter rises, however, are associated with non-contrastive contexts. That is, the trough has to be lowered or moved in order to avoid inconsistent stimuli, such as late peaks with a flat rise, although there is no basis for doing so in the production data. Recall that we encountered problems in labelling the exact position of the trough which might be one reason for the present paradoxon. We therefore propose two different manipulations to do justice to both possible interactions: In the experiment described in section 4.2.1, we moved the trough rightwards to get later peaks with the same slope, while in the experiment described in section 4.2.2, we lowered the trough for later peaks. Results from both manipulations are summarised and discussed in section 4.2.3.

4.2.1 Manipulation I (late peak = late trough)

For the first resynthesis experiment we assumed that later peaks are realised by corresponding later troughs. The stimuli were presented in contrastive and non-contrastive contexts.

Stimuli. Four versions of a rise were resynthesised in the four training and five target utterances presented in Table 4.1. This resulted in one version presumed to be non-contrastive (nc) and three versions presumed to be contrastive in varying degrees (see Figure 4.1).

For the experiments described in this section, we applied the following criteria for manipulating the stimuli: The *late-high* version was resynthesised by prolonging the rise of the *early-low* version (with the same slope). For the *late-low* versions, the f_0-minimum was moved to the right in order to have the same slope as the *late-high* version. The *early-high* version could not be manipulated with an earlier minimum because this would have moved the minimum into the pre-stressed syllable which sounded unnatural (an opinion that was shared by our colleagues). Therefore, we decided to increase the slope of the rise in the

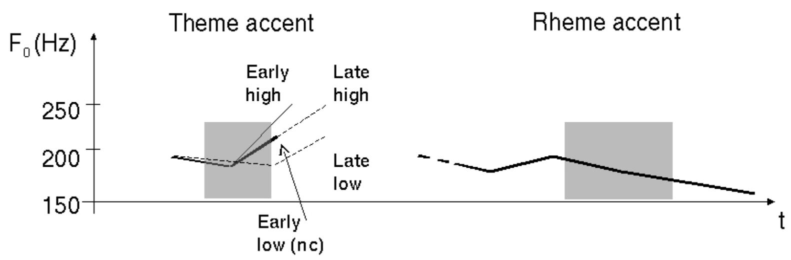


Figure 4.1: Schematic figure of the four resynthesised thematic rises of manipulation I. *Early-low* is presumed to be non-contrastive, all other versions are presumed to be contrastive. The f_0-excursion is 2.6st in the low peak versions and 5.2st in the high peak versions. The shaded areas mark the stressed syllables.

early-high version.

In more detail, as a starting point for resynthesis, a plausible non-contrastive realisation was created which had a rise to an *early-low* peak at 215 Hz at the start of the post-stressed unstressed syllable (see idealisation in Figure 4.1). The rise was preceded by a small trough, from 193Hz (start of the utterance) to 185Hz (in the middle of the stressed syllable)[7]. From that presumed non-contrastive version, the version with the *late-high* peak was created, prolonging the rise with the same slope to a peak of 250Hz. Obviously, the peak is reached at different points in the segmental structure (for details see Table B.1). From this *late-high* version, the peak was lowered, keeping the same slope and the same f_0 for L (i.e. the *position* of L is found in different segmental positions, see again Table B.1 for details). For the *early-high* version, the peak of the *early-low* version was raised to 250Hz, increasing the slope of the rise.

Since we were mainly interested in thematic rises, the rhematic contour of the utterance was always marked by a !H+L* (see Figure 4.1). This was the accent type most commonly found in the production data (see section 3.5.3). The middle of the stressed syllable was assigned a fundamental frequency of 165Hz. The prenuclear syllable (with the leading tone !H) was set at 190Hz (at the end of the respective syllable). In order to make the !H audible, the frequency of the syllable before that (= pre-prenuclear syllable) was realised with 180Hz. The boundary tone (L-%) was phonetically realised with a fundamental frequency of 155Hz at the last sonorant point of the utterance[8]. One might argue that the fundamental frequency of the rhematic accent should be adjusted to the peak height of the thematic accent

[7]These criteria can also be applied to the few stress-initial utterances in the training set.

[8]Note that one of the training utterances "In Indien konzentriert man sich auf Technologie" required an additional prenuclear peak on "konzentriert", for rhythmic reasons. The peak was realised on the syllable "iert" (205Hz) with the prenuclear syllable at 195Hz.

in order to achieve comparable prominence between the two accents. This was avoided because an increased thematic prominence is probably the complex phonetic way to express the contrast in this position. As Nolan (1995) noted, emphatic accents on certain constituents do not affect the scaling of following accents. In addition, there is too little knowledge and hence no operational criteria for adjusting the prominence relations without pursuing prior prominence adjustment experiments for each target utterance: "[...] perceived prominence is related in a complex way to the range of F0 values employed. [...] The findings make clear that prominence cannot be accounted for in terms of relations between F0 maxima only [...]" (Terken, 1991, p. 1768 and 1776).

Method. The stimuli were presented separately in contrastive and non-contrastive contexts. Participants were asked to rate the appropriateness of the intonation of the utterances on a seven-point scale (1 = very appropriate, 7 = very inappropriate).

Contrastive context: Ten subjects voluntarily participated in the first experiment in which the stimuli were presented in contrastive contexts. They were told to rate the appropriateness of the melody of the test utterances in the context of parallel constructions. Also, they were explicitly told to ignore the actual content of the sentences and any unnaturalness resulting from resynthesis. Sentences are similar to the following one (where "Marlene" was manipulated). The remaining examples are found in Appendix B.2.1:

Visual presentation:	Anna liebt das Meer, aber ...
	'*Anna loves the sea but ...*'
Auditory presentation:	Marlene fährt gern in die Berge.
	'*Marlene likes to go to the mountains.*'

The first half of the parallel construction was presented visually, followed by three dots to indicate the continuation. After clicking a loud-speaker symbol on the screen, the subjects heard a resynthesised continuation to the utterance. Five training examples were presented in fixed order to familiarise subjects with the speech stimuli and the task.

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 resynthesising the first part of the parallel construction, we avoid the as yet rather unexplored territory of inter-clause peak height relations.

Non-contrastive context: Nine subjects participated in the experiment, in which the appropriateness of the stimuli had to be judged in a non-contrastive context. As non-contrastive contexts we decided for the start of a story, assuming that contrastive versions would sound less appropriate. Subjects were asked to judge how well the intonation of a given stimulus qualifies for the start of a story[9].

Hypotheses. For the contrastive context, we hypothesised that the presumed contrastive versions (*early-high, late-high* and *late-low*) are judged as being more appropriate than the presumed non-contrastive version (*early-low*). Furthermore, we assumed that the versions with both higher *and* later peak would be rated better than versions with either a higher *or* a later peak[10]:

$$\boxed{\textit{late-high} > \textit{early-high} = \textit{late-low} > \textit{early-low}}$$

For non-contrastive contexts, we predicted a preference for the presumed non-contrastive version (*early-low*) over all the presumed contrastive versions. Note that because of the assumed trading relation, this preference pattern cannot be simply opposite to what is predicted for the contrastive context:

$$\boxed{\textit{early-low} > \textit{late-low} = \textit{early-high} = \textit{late-high}}$$

Results. Statistical analysis was performed using the non-parametric Friedman test and a Wilcoxon signed ranks test for post-hoc analysis. Surprisingly, subjects showed the same preference pattern in contrastive and non-contrastive contexts. In both contexts, listeners were able to significantly distinguish different versions (contrastive context: χ^2= 14.25, df=3, p=0.003, non-contrastive context: χ^2=18.6, df=3, $p < 0.001$). Post-hoc analysis showed that there was only a significant difference between *early-high* peak versions and all remainders. That is, subjects significantly preferred the versions with a steeper rise over ones with flatter rises:

$$\boxed{\textit{early-high} > \textit{late-high} = \textit{late-low} = \textit{early-low}}$$

Discussion. We first address the result that the context seemed to have no influence on the preference decisions of subjects. This is not in accordance with what we expected based on the production data, where the context was found to influence the prosodic realisation. This discrepancy between production and perception led us to search for alternative explanations for these unexpected results. From a phonetic point of view, the start of a story may be an unsuitable context, since initial elements are always marked more strongly than final ones. On the segmental level, onset consonants are articulated with a higher precision than coda consonants (Fougeron and Keating, 1996). There also seems to be some influence of a

[9]This “context” was rather unfortunate as will be discussed at the end of this section.

[10]Throughout the thesis “>” stands for “significantly preferred over” ($p < 0.05$), while “=” means “no significant difference”. In the results boxes, the displayed order of the versions is according to their respective mean ranks.

prosodic hierarchy, with paragraph-initial consonants being more precisely articulated than phrase-initial ones. Prosodically, according to Grosz and Hirschberg (1992), the start of a new paragraph is accompanied by a pitch reset and higher f_0-range (cf. also Wichman et al., 2000; Shriberg et al., 1998). F_0-differences have been shown to be important for an automatic sentence and topic segmentation Shriberg et al. (1998).

In addition to these phonetic peculiarities, the start of a story is not a sufficiently defined context. Crain and Steedman (1985) pointed out that the presentation "out-of-context" is not necessarily a null-context but rather an uncontrolled one. As a consequence, subjects might "imagine" — or in a more discourse-oriented term "accommodate" — all kinds of contexts that the experimenter cannot control.

The present experiment revealed that subjects only accepted versions with a steeper rise (in both contexts). This importance of the slope of an accentual rise on perception was also found in Bartels and Kingston (1994) who investigated rhematic contrast. Since neither peak height nor peak position seemed to influence subjects' appropriateness judgements (although they clearly differed in the *production* of thematic elements in different contexts), we assumed that the resynthesised versions used might be somewhat unnatural. Because of the difficulty in determining how to resynthesise the trough before the accentual rise, it might have been wrong to implement a later peak by delaying the trough. In section 4.2.2 we therefore describe an experiment with an alternative manipulation where later peaks are accompanied by lower troughs.

4.2.2 Manipulation II (late peak = low trough)

The experiment with contrastive contexts was replicated with different resynthesis-criteria. This was motivated by the results of the previous experiment where subjects based their decision solely on the slope of the rise so that peak height and peak alignment became irrelevant factors. For phonetic reasons, the slope could not reasonably controlled in the previous experiment. In the manipulation described and investigated here, higher peaks are accompanied by steeper slopes.

Stimuli. The *early-low* versions in this experiment were the same as in the previous experiment (rise to an *early-low* peak at 215 Hz at the start of the first post-stressed syllable, see Figure 4.2). From that realisation the peak was delayed until the end of the word. This time, the later peak was realised with a lower trough (170Hz instead of 185Hz) to keep the slope constant. From the *late-low* version, the peak was raised to 250Hz for the *late-high* version, which resulted in a steeper slope. This is in line with the findings from the production data. The *early-high* version was derived from the *early-low* version, simply with a higher peak (and a steeper slope respectively). Similar to the previous study, the rhematic contour was always marked by a !H+L* (see Figure 4.2)

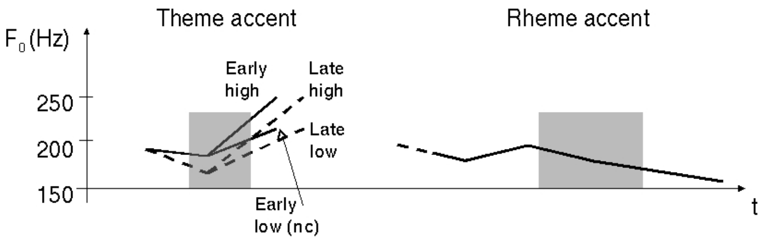


Figure 4.2: Schematic figure of the four resynthesised thematic rises of manipulation II. As before, *early-low* is presumed to be non-contrastive, all other versions are presumed to be contrastive. The f_0-excursions in the different versions are 2.6st for *early-low*, 4.0st for *late-low*, 5.2st for *early-high*, and 6.6st for *late-high*. The shaded areas mark the stressed syllables.

Method. Sixteen subjects were asked to rate the appropriateness of the *intonation* of the respective continuation with respect to the first part of a parallel construction using a seven-point scale.

Hypotheses. The present hypotheses are comparable to the ones of the previous study: It was predicted that listeners prefer all contrastive versions (*early-high*, *late-high*, and *late-low*) over the non-contrastive one (*early-low*). Furthermore, we assumed that a stronger marking of contrast (by both a later *and* higher peak) is preferred over weaker cues for contrast (later *or* higher peak). Thus we assumed the following preference ranking:

$$\boxed{\textit{late-high} > \textit{early-high} = \textit{late-low} > \textit{early-low}}$$

Results. Normalised data were statistically analysed using the non-parametric Friedman test, together with a Wilcoxon signed ranks test for post-hoc analysis. Observing subjects' responses during the experiment indicated that there are different rating behaviours among subjects. More specifically, we found two groups of subjects, one group with twelve subjects who preferred the intonationally better marked contours (later and higher peaks) and one group with only four subjects who showed no clear preference or — if any — a preference in the opposite direction. We therefore formed two groups and calculated an ANOVA with `rating`[11] as within-subject factor and `group` as between-subjects factor. We found a significant interaction between `group` and `rating` $F(3,234)=5.09$, $p=0.02$. That is, there were indeed two different rating "behaviours" or "strategies", so we analyse and discuss the two groups separately hereafter.

[11] For easier readability, factors in ANOVAs are printed in `true type` throughout the thesis.

For both groups, the non-parametric Friedman test was calculated. The four subjects in the smaller group preferred all versions alike. For the remaining twelve subjects, their rating differed significantly in the four conditions: χ^2=28, df=3, $p < 0.001$. A post-hoc analysis (Wilcoxon signed ranks test) revealed that all contrastive versions differed significantly from the non-contrastive version: $p < 0.001$, $p < 0.001$, and p=0.02, respectively. Furthermore, all conditions, except for the two stimuli-versions with a high peak, were judged significantly differently from each other, suggesting a "categorised continuum". However, the hypothesised preference grouping (preference for *late-high* over both *early-high* and *late-low*) was not observed. Instead, the two versions with a steeper slope were grouped together and rated best. This is summarised in the following results box:

$$\boxed{\textit{early-high} = \textit{late-high} > \textit{late-low} > \textit{early-low}}$$

Discussion. This discussion only focuses on the group of twelve listeners who showed a distinctive preference pattern. They were able to perceive a difference between the non-contrastive version (*early-low*) and all contrastive versions. However, the hypothesised preference for the strongly marked contrast (*late-high*) was not observed. Rather, listeners did not judge the two versions with a high peak differently, rather, both high peak versions were judged as being most appropriate in the given contrastive context. That is, if the peak was high, differences in peak alignment were either not perceived or played no role in the present context. This might be explained by a ceiling effect, i.e. peak height alone might have been a sufficient cue to signal contrast. Because of this possible ceiling effect we cannot answer the question whether peak height and peak alignment are additive factors or not.

The resynthesised versions with a *late-low* peak were judged worse than both high peak versions but still differed significantly from the non-contrastive one (*early-low*). In other words, it appears that peak alignment (or the accompanying lower trough) was crucial only for the interpretation of low peaks. The importance of peak height on perception is in line with other studies. Bartels and Kingston (1994), for instance, summarise that "peak height was found to be by far the most salient cue to contrastiveness" (p. 1). It is worth noticing, however, that the perceptual importance of the scaling of accents has been largely ignored in autosegmental tonal categories (e.g. Pierrehumbert, 1980; Beckman and Ayers, 1997) and is acknowledged only in perceptually-oriented annotation schemes, like the British School (e.g. Crystal, 1969). An exception worth noting was the attempt to combine special features (e.g. [delayed peak], [downstep], and [raised peak]) with standard autosegmental theory as proposed by Ladd (1983).

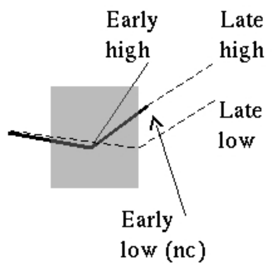
The results indicate that there is a trading relation between peak height and peak alignment with a stronger impact of peak height. Comparing this to the realisations of all utterance pairs in production (i.e. an utterance produced by one speaker in two different contexts) shows that perception almost mirrors production. In production, 70.15% of the pairs exhibited peak height differences, and only 29.85% peak alignment differences (see Figure 3.7 in the previous chapter). Surprisingly, the combination of higher and later peaks which was most common in production (47.8%) was not preferred most in perception, which might be due to a ceiling effect of pitch height.

Alternatively, one could consider overall f_0-excursion instead of peak height as underlying factor that influences the appropriateness ratings. In contrast to absolute peak height, f_0-excursion would help us to explain the special intermediate status of the *late-low* version in manipulation II. Its f_0-excursion lies between the *early-low* and *early-high* version as did its preference score. That the versions with a high peak were judged as equally appropriate — despite their different f_0-excursions — might again be attributed to the already indicated ceiling effect.

4.2.3 Evaluation and Conclusion

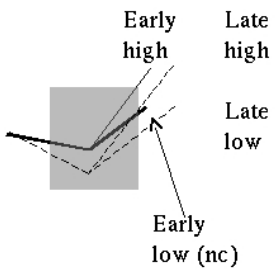
In the last section we investigated the influence of peak height and peak alignment on perception. Since the status of the trough before the rise remained rather unclear in the production data, we proposed two different manipulations. In the experiments described in 4.2.1, we resynthesised late peaks with an accompanying late trough and in the experiments discussed in 4.2.2, late peaks were realised by lower troughs, resulting in the two sets of stimulus versions summarised in Figure 4.3.

Manipulation I (section 4.2.1)



early-high > late-high = late-low = early-low

Manipulation II (section 4.2.2)



early-high = late-high > late-low > early-low

Figure 4.3: The two prosodic implementations together with the preference patterns gained from the contrastive condition.

In this section we link the results from both experiments in contrastive contexts and discuss their implications. Since we believe that the context for the non-contrastive condition was not sufficiently defined, it would only distract from the main argument to discuss these results again here. We shall come back to this issue in section 4.3.1.

The preference pattern in contrastive contexts have shown that alignment is a rather weak cue in the perception of contrastive themes. In manipulation I (left part of Figure 4.3), versions with later peaks were not judged to be appropriate at all. In manipulation II (right part of Figure 4.3), alignment differences seemed to have an impact only when the peak was low. Therefore, we concluded that peak height differences are a far stronger cue to thematic contrast marking than alignment differences. This result contrasts with the findings of a prominence-based perception experiment by Kohler and Gartenberg (1991) who reported a strong effect of peak position. They manipulated both peak position (early, medial, and late peaks with constant slope)[12] as well as five different peak heights (100-180Hz in 20Hz steps) of nuclear accents in German. They found that "medial peaks are more prominent than either early or late peaks and need a substantial decrease in height compared with the other two types to equalise prominence" (Kohler and Gartenberg, 1991, p. 239). In their study, the position of the peak appears to be a rather robust cue that only very strong differences in peak height could compensate. In German, the distinction between early and medial peaks in nuclear accents was shown to be categorically distinct (Kohler, 1991b). Our results suggest that the perception of *prenuclear* accents is less sensitive to differences in peak position. Or, taking into accout that the peak is only found in the post-stressed syllable in our data, perceptual differences between late and very late peaks are not very clear. In Kohler's data also, medial and late peak were not categorically distinct.

The picture still remains contradictory when we discuss the results of both manipulations with respect to the factors peak height and extent of f_0-excursion: In manipulation I, *early-high* is preferred over *late-high* although they have the same peak height and f_0-excursion. In manipulation II, however, these versions form one group that was rated best although *late-high* has a greater f_0-excursion than *early-high*. This finding would make slope the most important perceptual factor. This is unexpected because the slope of the rise was not among those variables that differed significantly in the t-tests performed on the production data. Note, however, that we adopted a fairly conservative significance level to account for the large number of tests performed. This, in turn, increased the probability of not finding an existing difference. In this respect, our interpretation of the results of the production data might thus have been slightly misleading. Interestingly, the slope of accentual rises has recently been classified as epiphenomenon by Arvaniti et al. (1998). For Greek, they found stable segmental anchor points for both H and L. Because of these contradictory findings, we calculated the correlation between peak height and slope in the production data for which we found a significant correlation ($r=0.82$, $p < 0.001$). This correlation might indicate that the *late-high* versions of manipulation I were not adequately modelled (i.e. their rise is not steep enough for the high peak). The correlation indicates that a steeper slope is thus concomitant with higher peaks[13]. That slope cannot be the single most important factor becomes clear in the low peak versions of manipulation II. There, the simple explanation "same slope – same preference" is not applicable: the versions *late-low* and *early-low* have

[12]These accents can be mapped onto the autosegmental categories H+L*, H*, and L+H*, respectively.

[13]Many studies that investigate different peak heights actually increase the slope of the rise (cf. Chen, 2003; House, 2003).

the same slope but *late-low* was preferred over *early-low*. That is, either the difference in peak alignment or the extent of f_0-excursion seems to play a role as well. The alignment of the peak to the segmental structure, however, is less important, which might indicate a free variation[14]. This free variation could be due to a dependence of alignment on some other factor(s). One possible factor that can easily influence alignment — and which we have neglected so far — is duration[15]. And indeed, there is a significant correlation between the temporal distance of the peak from the start of the stressed vowel (al(H*,V0)) and syllable duration ($r=0.54$, $p < 0.001$) as well as word duration ($r=0.43$, $p < 0.001$). That is, the peak is later when segmental durations are increased. This is in accordance with converging evidence that the peak is delayed when the duration of the stressed syllable is increased, but only when it is not lengthened by an upcoming prosodic boundary (cf. Steele, 1986; Silverman and Pierrehumbert, 1990; Prieto et al., 1995; Kim, 1999; Grimm, 1997). Again, these results are based on nuclear accents (in different languages) and do not allow a 1:1 comparison to our data. Apart from a causal relation between these two variables, segmental duration is likely, independently, to be an important contributing factor to the perception of prominence, or more specifically to the perception of thematic contrast.

There is a slight suspicion that the presumed non-contrastive versions were not preferred because these *early-low* versions were sufficiently long so that they could be judged as prominent. Therefore, our next step was to explore the influence of segmental duration and peak height (which, in our resynthesis, was tantamount to slope) on the perception in both contexts (section 4.3). We do not think it is very useful to continue to use stimuli that are no longer in accordance with the latest level of knowledge. Therefore we abstain from testing the present stimuli in non-contrastive contexts. We are aware of the fact that investigating only the perception in contrastive contexts is not sufficient to draw conclusions about the perception of contrast. The results of the previous experiment in non-contrastive context have shown that we need to have a solid basis for rejecting the null-hypothesis that context has no influence on the preferred realisation. An explanation in favour of the null-hypothesis is that subjects always prefer the intonationally stronger marked realisations over less marked ones perhaps because these sound more interesting and more lively than versions with flat f_0-rises. To make semantically valid claims about the perception of contrast, however, we need to find an influence of contrast on preference judgements. The stimuli of the following experiments will again be evaluated in contrastive and non-contrastive contexts.

[14]Manipulation II revealed a less preferred alternative, namely a later peak that is realised by means of a lowered trough. Unlike the high peak versions, this one is *not* accompanied by a steeper slope.

[15]Up to now, this factor has not been manipulated because we wanted to avoid complicating the experimental setup. Presenting all versions with both a longer and a shorter segmental duration would have doubled the amount of data.

4.3 Influence of Peak Height and Segmental Duration

In contrastive contexts, peak height seemed to be perceptually more important than peak alignment. The task we used for non-contrastive contexts (start of a story), however, was neither well suited to define a strictly non-contrastive context, nor was it phonetically appropriate. In the first section (4.3.1), we therefore discuss a more appropriate non-contrastive context.

A further change compared to the previously described method is that the stimuli in the experiments hereafter vary in both segmental duration and peak height. The duration of the stressed syllable could be an important contributing factor to the perception of contrast, an issue that we have not accounted for so far. Note that what we treated as genuine alignment effects in the production study could actually have been caused by durational differences in contrastive and non-contrastive contexts. Consequently, in the following experiments, we manipulated the variables peak height and segmental duration rather than peak height and peak alignment. An unexpected outcome of the pilot experiment described in section 4.3.1 also gave us the opportunity to learn more about the, as yet neglected, rheme-accent on perception.

4.3.1 Pilot Studies for Non-Contrastive Contexts

The previous experiment with a non-contrastive context has shown that it is important to maximally restrict alternative interpretations (which is more difficult in non-contrastive contexts than in contrastive ones). For this reason we decided on a question-answer methodology (Q-A) with focus-inducing questions as precontexts. This allowed us to control the information structure of the target utterances in the best possible way.

Method. To establish a non-contrastive context, focus-inducing questions established the theme-rheme structure. The target utterances constitute the answers to these context-questions. The questions, which were supposed to be requests for an explanation, were recorded by the author. The constituent that would end up as the pre-verbal thematic material of the answer was intonationally marked as given in the corresponding question (by deaccenting it). The questions' boundary tone was always marked by a rising accent. Below, an example is given of one of the mini-dialogues used (the remaining stimuli are summarised in Table B.2 in the Appendix). Capital letters mark the position of the nuclear pitch accent, the focus-exponent. Only the unambiguously definable information structure parts are annotated. Thematic parts are labelled with $[_\theta \ldots]$, rhematic, new parts with $[_\rho \ldots]$[16].

[16] The notation to mark θ for thematic and ρ for rhematic elements is taken from Steedman (2000). In Vallduví and Engdahl (1996), these are referred to as *Link* and *Focus*, respectively.

Question:	Was [$_\rho$ MACHT] [$_\theta$ Marlene] denn gerne?
	'What does [$_\theta$ Marlene] [$_\rho$ like to DO*]?'*
Answer:	[$_\theta$ Marlene] [$_\rho$ fährt gern in die BERGE].
	'[$_\theta$ Marlene likes] [$_\rho$ to go to the MOUNTAINS*].'*

Context questions and stimuli were presented in one sound file, with a fixed pause of 500ms separating them. Like in the previous experiments, subjects had to rate the appropriateness of the answer's intonation with respect to the context-question on a seven-point scale. They were allowed to play the mini-dialogues as often as they wanted. We told them that the task was difficult and that they should rely on their intuition. The mini-dialogues were not presented visually to reduce the influence of silent prosody when reading (cf. Fodor, 2002). The influence of silent prosody probably increases when subjects do not decide intuitively but take time to read over the dialogue again. Then, subjects might not judge the intonation of the actual stimulus they heard but base their decision on their inner representation of it, which is likely to be guided by the written presentation.

Stimuli. In this and the subsequent experiments, both the segmental duration of the stimuli and the f_0-excursion were manipulated. This resulted in four resynthesised versions. Stylisations are shown in Figure 4.4.

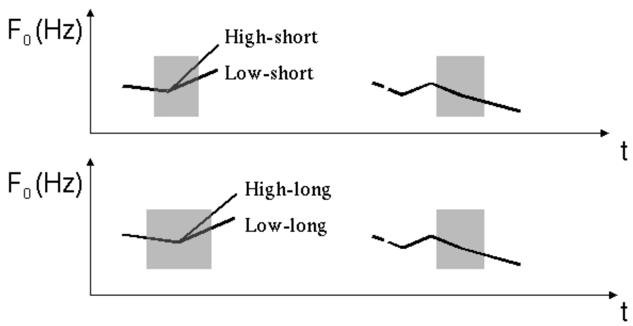


Figure 4.4: Manipulation of peak height and segmental duration. Versions with shorter duration are depicted in the upper part, versions with longer durations in the lower part.

For the manipulation of f_0-movement we used those versions of manipulation II that yielded best and worst appropriateness values in the study with contrastive context (*early-high* and *early-low* in Figure 4.2). The duration of the whole preverbal constituent (prosodic word, including prenominal articles or prepositions) of the existing stimuli was shortened to 85% according to the results of the production study[17]. Since slope seemed to be important for perception (which was also emphasised by Bartels and Kingston, 1994; Xu, 2002), it

[17] We thereby avoided new recordings of our original speaker, which might have resulted in different speech rate and voice quality characteristics. Duration manipulation was carried out using Praat.

was held constant across duration conditions. Similarly, the alignment of the f_0-minimum (L) was kept constant, according to various studies that found a consistent alignment of L with the segmental structure (e.g. Caspers and van Heuven, 1993; Prieto et al., 1995). This resulted in a slightly later peak in conditions with shorter constituents[18] as shown in Figure 4.4.

Hypotheses. According to the results of the production study we predicted that less marked themes with a shorter duration and a lower peak height should be more appropriate in non-contrastive contexts than stronger marked versions (longer duration or higher peak). As both *low-long* and *high-short* version contain just one cue to prominence we assumed that they would be judged as equally appropriate:

$$\boxed{\textit{low-short} > \textit{low-long} = \textit{high-short} > \textit{high-long}}$$

Results. There was no significant preference for any of the resynthesised versions, i.e. all versions were judged as equally appropriate. Some of the subjects reported that they had been highly irritated by the low rhematic accent (!H+L*) on the focus-exponent (that would be "Berge" in the present example, see page 106). Although this was the accent type most often found in the production data, they complained that the "main accent" was not "strong enough" and that the utterance was "wrongly accented".

Discussion. The finding that differences in prosodic theme-marking do not influence subjects' rating in non-contrastive context was already made in the previous rating experiment. This time, there was some indication that subjects encountered difficulties because of the rheme accent type used. The "wrong" rheme accent may have decreased subjects' ability to perform the task. It might be that they were so distracted by the low rhematic accent that they did not perceive any difference in the thematic part. Or, the wrong accent on the new information was so severe that nothing in the thematic part could improve the acceptability.

Although we found a majority of low rheme accents in the production data, many subjects complained that the "main accent" (i.e. the nuclear, rhematic accent) was not strong enough and hence did not "fit the question". This impression that early peaks (i.e. low accents or !H+L* or H+!H* in autosegmental terms) are "not strong enough" has been experimentally shown by Kohler and Gartenberg (1991) who manipulated both peak height and peak alignment in single-accent utterances in German. Subjects were presented pairs of

[18]Note that Arvaniti et al. (1998) found a stable anchor for H in addition to the stable anchor for L, which reduces the slope of the rise to an epiphenomenon. A main difference between their study and ours is, however, that they aimed at exploring the factors that affected the position of H (in words with antepenultimate stress), while we are more generally interested in a presumed meaningful difference in two different sorts of rising accents. This variation (within one intonational category) has to be signalled appropriately.

such manipulated utterances and had to decide which of the versions was more prominent. Ignoring a very strong order effect, subjects rated medial peaks as being more prominent than early or late peaks (when the peak height was the same). It turned out that "early peaks must be over 40Hz higher to be perceived by the majority of listeners as more prominent [than medial peaks]." (Kohler and Gartenberg, 1991, p. 234). A similar observation was made by Ayers (1995), who therefore concluded that "nuclear accent is not a completely uniform category in terms of prominence" (p. 663).

The semantic interpretation of early peaks has been described (and experimentally evaluated) as expressing an established fact rather than conveying new information (cf. Kohler, 1991b). Despite the prosodic "weakness" of these accents, this interpretation would fit the situation expressed in our dialogues where a question is then answered "matter-of-factly" instead of politely supplying new information. However, subjects apparently could not adopt this interpretation.

This asymmetry between production and perception might be explained by the two different speech styles investigated: in the production experiments we used read speech while the question-answer pairs used in the perception experiments are closer to spontaneous speech. It might be that the high proportion of low accents in the production data is a slight artefact of the reading task, i.e. they might be much rarer in spontaneous speech. Alternatively, the mini-dialogues could have been interpreted as too formal to accept the less polite "established-fact" interpretation. This result raises several further research questions from which we will try to address the following: If non-contrastive contexts force a special kind of rheme accent, does this apply to contrastive contexts as well? This could be expected from the distribution of rheme accents in the production data. For contrastive contexts, there were significantly more low rheme accents than high ones (χ^2=13.44, df=1, $p < 0.01$).

In the following main experiment, we replicated the pilot experiment described here with high rheme accents (!H* in autosegmental terms) and a more elaborate precontext. The experiment was performed in contrastive and non-contrastive contexts. The contrastive context condition will help to answer the question whether a different rheme accent type influences subjects' performance to the same extent as it did in non-contrastive contexts.

4.3.2 Experiment

Method. Based on previous experience the stimuli were resynthesised with high rheme accents (205Hz on the nuclear syllable, 180Hz on the preceding syllable) to make the rhematic part more appropriate to the question. As described above, stimuli varied in both f_0-movement and segmental duration. A further refinement was the use of magnitude estimation instead of a seven-point scale. Magnitude estimation is a psychophysical method originally used by Stevens (1957) for the scaling of loudness. This change in rating-paradigm was mainly motivated by questions from subjects who were insecure to which "reference" they should compare, i.e. some subjects felt uncomfortable with an absolute rating based on a fixed scale. Furthermore, there were subjects who mainly used the middle point of the scale although we had encouraged them to be "confident" in their ratings. Magnitude estimation has the advantage that subjects can freely choose the perceptual detail they wish to express. Furthermore, they are able to compare the current realisation to the previous one and make a relative judgement. Such relative judgements seem to be easier when the categories to judge are not well established. At school, for instance, everyone has an idea what grade "4" means compared to grade "3" but in the current tasks, it is not quite clear how appropriate "4" is compared to "3". Magnitude estimation has recently been applied in a variety of linguistic rating tasks and has proved a highly reliable and highly replicable method to elicit consistent judgements (see results of cross-modality matching by Bard et al. (1996), in which judgements were validated using different response modalities, such as line lengths and numeric values). Not only is the magnitude estimation technique suitable for assessing the quality and appropriateness of speech stimuli (cf. Pavlovic et al., 1990), the results of this method are also comparable with ratings on other subjective scales, such as pair-wise comparisons (cf. Pavlovic et al., 1989) or ratings on ordinal scales (cf. Keller, 2000).

To familiarise subjects with the magnitude estimation task they were first asked to judge the length of eight lines. It was pointed out to them that a line twice as long as the previous one should receive a rating twice as high as the previous one. Afterwards they were told that this method was now being used for judging the appropriateness of speech data. Since it is hardly possible to remember the speech pattern of a reference utterance ("modulus"), we told them already in the line-rating task to make their rating with respect to the directly previous stimulus. Theoretically, this kind of rating does not differ from a rating with respect to a reference item. Practically, however, the errors accumulate slightly when comparing to the previous stimulus.

Non-contrastive context: To strengthen the intended non-contrastive interpretation, the precontext was extended by visually presenting two or three sentences with the same topic as the theme of the answer (= target utterance). It was suggested to subjects that this could be interpreted as a dialogue with the precontext said by the same speaker who subsequently provides the answer. Again, unambiguous thematic elements are marked by $[_\theta \ldots]$, rhematic elements by $[_\rho \ldots]$. For the full set of experimental contexts see Appendix B.2.3.

Visual Context:	Ich hab keine Ahnung, was ich Marlene zum Geburtstag schenken soll. Hast du vielleicht ne Idee?
	'I really don't know what I could buy as a birthday present for Marlene. Do you have an idea?'
Auditory Question:	Was [$_\rho$ MACHT] [$_\theta$ Marlene] denn gerne?
	'What does [$_\theta$ Marlene] [$_\rho$ like to DO*]?'*
Auditory Answer:	[$_\theta$ Marlene] [$_\rho$ fährt gern in die BERGE].
	'[$_\theta$ Marlene] likes [$_\rho$ to go to the MOUNTAINS*].'*

Contrastive context: For the contrastive context, we used the parallel constructions already described in section 4.2.1.

There were two groups of subjects, one that was presented the non-contrastive context first and one that was presented the contrastive context first (23 subjects altogether).

Hypotheses. Obviously, the hypotheses for non-contrastive and contrastive contexts are direct opposites.

From the results of the production study we hypothesised that the less marked themes should be more appropriate in non-contrastive contexts. Since the *low-long* and *high-short* versions contain just one cue to prominence we assumed that they would be grouped together:

$$\boxed{\textit{low-short} > \textit{low-long} = \textit{high-short} > \textit{high-long}}$$

For the contrastive context we expected the opposite pattern, i.e. that stronger marked versions (with a longer duration or a higher peak) would be more appropriate. Again, we assumed that f_0-movement and duration could be additive factors[19]:

$$\boxed{\textit{high-long} > \textit{high-short} = \textit{low-long} > \textit{low-short}}$$

Results. Magnitude estimation values were log transformed ($\log_{10}$) and normalised. In contrast to the ratings in the previous studies, these are scalar data which can be analysed using parametric tests. To investigate the influence of presentation order (indicating which contrast condition was rated first), we calculated an ANOVA with `rise type` as within-subject factor (with the four conditions *low-short, low-long, high-short* and *high-long*) and `order` as between-subject factor for contrastive and non-contrastive contexts. In non-contrastive contexts, there was a significant interaction between `order` and `rise type`: $F(3,276)=3.57$, $p=0.015$. That is, the non-contrastive ratings of subjects that were in the

[19]Note that the discrepancy between the high proportion of low rheme accents in contrastive contexts in production and the resynthesised high rheme accents in the stimulus utterances could lead to processing difficulties.

group that performed the non-contrastive task first differed significantly from the respective ratings of the subjects in the other group that rated the contrastive contexts first. It is rather likely that for the one group of listeners, the acceptability judgements in non-contrastive contexts were influenced by the preceding ratings in contrastive contexts. Therefore, for non-contrastive contexts, we only analysed the results of the group of subjects that first performed the non-contrastive task (15 subjects). In the contrastive contexts, there was no significant interaction between `order` and `rise type`, so ratings were pooled (23 subjects).

Non-contrastive context: A repeated measures ANOVA with `peak height` and `duration` as within-subject factors was calculated. For the 15 subjects that performed the non-contrastive context-task first, we found no effect of peak height or of duration and no interaction[20]. This result of the main group is comparable to the result of the pilot test, in which stimuli had a low rheme-accent and a simple Q-A task was used.

Contrastive context: Listeners encountered no difficulties because of the high rhematic accent. There was a significant effect of `peak height` ($F(1,95) = 18.13$, $p < 0.01$), but no effect of `duration` and no interaction[21]. Subjects had a clear preference for versions with a high peak over versions with a low peak (as expected). Duration had no effect on preference ratings:

$$\boxed{\textit{high-long} = \textit{high-short} > \textit{low-long} = \textit{low-short}}$$

Discussion. As opposed to non-contrastive contexts, a different kind of rheme marking in contrastive contexts was not commented on. That is, finding the right kind of rheme accent was crucial only in non-contrastive contexts, not in contrastive ones. This might indicate a free variation in rheme accent type for contrastive contexts only. Alternatively, this difference might be partly explained by the *visual* presentation of the parallel constructions, which allowed more interpretational (and hence intonational) freedom than the rigid questions used to induce non-contrastive contexts. The non-contrastive interpretation for the theme was established by a question that focused on the new, rhematic part. It is interesting to compare this to the production data where we also found a significant difference in the distribution of rheme-accents but in reverse order: in contrastive contexts, there

[20]For the remaining eight subjects whose performance might have been influenced by the previous contrastive context task, subjects preferred the *high-short* version over all other versions. Since we do not know what effect the contrastive-context task had on the performance of these subjects, we do not further discuss the results of this group.

[21]For those who suspect that this result was merely caused by analysing more subjects than were analysed in the non-contrastive context we calculated ANOVAs with equally sized samples as in the non-contrastive context. In all those analyses, we found the same preference patterns and comparable significances.

were significantly more low rheme accents, while high and low rheme accents were equally distributed in non-contrastive contexts.

In the contrastive contexts, acceptability was solely based on f_0-movement, not on duration. This is not really surprising given that f_0-excursion is generally not strongly influenced by different speaking rates (Caspers and van Heuven, 1993). That is, segmental duration can be shortened without affecting the extent of f_0-movement (and — in our case — the contrastive interpretation). In terms of contrastive theme marking, peak height appears to be the primary cue, followed by peak position; durational differences seem negligible.

Again we found that listeners judged all versions equally appropriate in the neutral, non-contrastive contexts. Based on the scientific question, we took it for granted that subjects based their judgements on the subtle variations in the thematic part that was neither conversationally nor information-structurally salient. As Cutler and Fodor (1979) have shown, however, reaction-times in a phoneme-monitoring task were significantly lower when the respective phonemes were part of a focused word than when they formed a thematic, given word. If we transfer these results to the present experiment, subjects might have had difficulties in perceiving the resynthesised differences in non-contrastive contexts. It might be suspected that the resynthesised differences did not exceed the threshold of consciousness so that subjects mainly judged the rhematic accent. This would better explain the complaints that arose from the presentation of an "inappropriate" rhematic accent in the non-contrastive context in the pilot experiment. If there is indeed a higher perceptual threshold for material that is not "in focus", our thematic rises will have to be more distinct in order to be judged as inappropriate. In contrastive contexts, however, the parallel construction highlights both information-units because there is a contrast in the theme *and* in the rheme. Thematic contrast can thus be seen as a focusing operation, a view also taken by van Hoof (2003)[22].

The finding that even intonationally strongly marked theme accents were accepted in contrastive contexts can be related to the experiments conducted by Nooteboom and Kruyt (1987). In a rather different experiment, where they investigated the accentuation and deaccentuation of different information units, they reported that the deaccentuation of new (rhematic) elements was hardly accepted, while accenting given (thematic) information was accepted quite often. How can the acceptability of "unnecessary" intonational marking in their and our experiments be explained? One illustration is found in Andreeva and Barry (1999). Similar to our findings they reported that in clearly marked neutral contexts (in their experiment declaratives), marked forms (in their experiment questions) were accepted but seen as carrying additional meaning (e.g. emphasis). Since the opposite does not apply,

[22] This analysis is not in line with the conclusion of Terken (1985), who argues on the basis of his experiments that "the presence of an accent instructs the listener to concentrate on the accurate analysis of the speech signal [...]" (Terken, 1985, p. 103). In our case it rather seems that the *context* — rather than the accent distribution — determines listeners' attention.

non-contrastive contexts can also be seen as prosodically under-specified so that they can give room for alternative functions of prosody (e.g. impatience, strong opinion). Contrastive contexts, on the other hand, appear to strongly constrain the prosodic realisation.

A further explanation is that subjects "repair" the context to accommodate for the perceived contrast (cf. Thomason, 1990). This assumes that intended contrastive versions are accepted in non-contrastive contexts because listeners manage to construct a context containing alternative items. In other words, intended non-contrastive versions are perceived as actually contrastive, and the non-contrastive context is converted into a contrastive one. This is similar to how semantic theories formalise (thematic and rhematic) accents (cf. Rooth, 1992; Büring, 1997; Steedman, 2000).

4.4 Alternative Functions of Prosody

In this section we describe a pilot study that was designed to shed light onto the unexpected (but repeated) finding that subjects preferred a large range of stimuli in non-contrastive contexts. In the previous section we devised two hypotheses: a) that non-contrastive contexts are prosodically underspecified so that prosody is free to convey other functions apart from information structure and contrast and b) that subjects accommodate a possible contrast which is not linguistically encoded. In the rating experiments described in chapter 4, subjects were asked to judge the appropriateness of one particular (resynthesised) stimulus utterance in a given context. In the experiment described here, we explored whether subjects are more likely to decide for the presumed non-contrastive version if they were to decide between two versions, one presumed to be non-contrastive and one presumed to be contrastive. After having made a choice, subjects were further asked to describe why the utterance they selected was more appropriate than the other one. We were aware of the unreliability of introspective data (Nisbet and Wilson, 1977; Ericcson and Simon, 1980) but hoped to find some indications to the interpretation of phonetically different versions in non-contrastive contexts.

Stimuli. As stimuli we used the same sentences as in the afore mentioned rating experiments. For each of the five sentences (see Appendix B.2.3) we chose the two versions that were judged as most and least appropriate in contrastive contexts in the previous experiment (high-long and low-short, as described in section 4.3.2). As familiarisation stimuli we used two of the training utterances from the previous experiments.

Method. Ten subjects voluntarily participated in the experiment. They were seated in front of a laptop and could listen to the stimuli via headphones at a comfortable loudness. The experiment was designed as a web-interface, see Figure 4.5 for a screen-shot. Subjects were presented the extended pre-contexts used before to ensure a non-contrastive interpretation of the thematic material. After clicking on the loudspeaker symbols labelled with a) and b) they could listen to the focus-inducing question and one of the two manipulated versions. The question-answer pairs could be played as often as necessary. The stimulus sentences were presented in random order. Also, the presumed contrastive and non-contrastive versions were randomly assigned to either a) or b) to avoid a bias.

Results. Subjects significantly preferred presumed non-contrastive versions (prenuclear accents with *low-short* rises) over presumed contrastive ones (prenuclear accents with *high-long* rises), χ^2=9.7, df=1, $p < 0.001$. That is, in direct comparison, utterances with phonetically less marked versions were judged to be more appropriate than strongly marked ones. Reasons subjects gave for this kind of preference can be summarised as information structural: position of the main prominence and appropriateness to the question. In addition, they often rejected the stronger-marked versions because these seemed to establish a contrast to the thematic element, which they thought was inappropriate.

In cases in which subjects preferred the presumed less appropriate version, the following reasons were given: the version sounded more "agreeable", more appropriate with respect to the literal semantic meaning (i.e. "resigned", "good-tempered", "nerved"), "more enthusiastic", less "businesslike", had the "emphasis on most important word" (i.e. subject), and was more "explaining" (instead of reproachful).

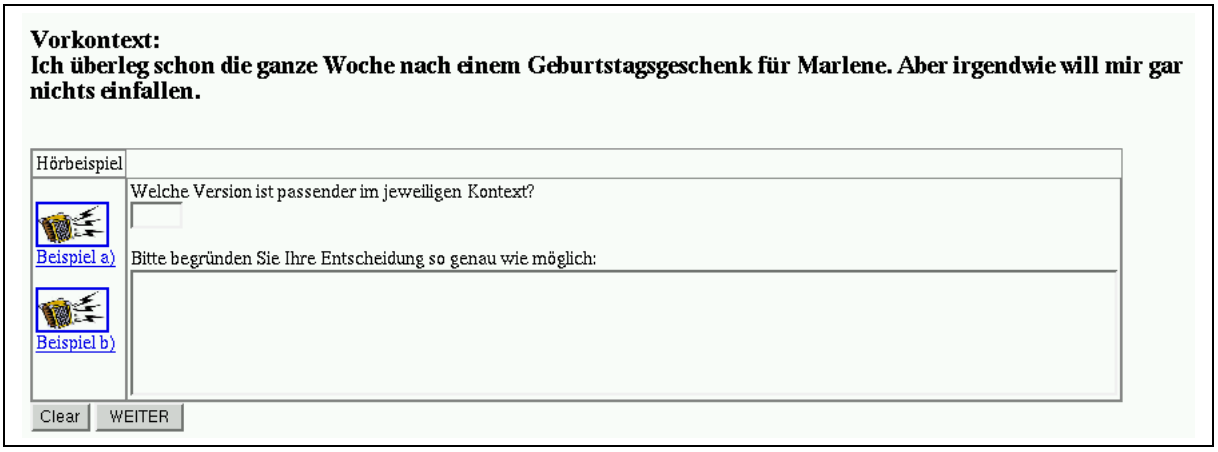


Figure 4.5: Interface for our experiment to test alternative functions of prosody. Participants could hear the two question-answer pairs after clicking on the loudspeaker symbols. They had to decide which of the two versions (a or b) was most appropriate in the context of the focus-inducing question and had to explain their choice.

Discussion. The results of this pilot experiment have shown that there are different "strategies" to process and interpret the utterances' prosody. One such strategy could be described as information structural because subjects' decisions relied mainly on the appropriateness of the prominence distribution in a given context. Another strategy was that subjects evaluated rather the presumed attitudinal state of the speaker or its effect on the listener. Finally, some subjects seemed to prefer versions in which the topic was sufficiently marked. That is, for them, a strongly marked preverbal constituent was not interpreted as contrastive (while it was for most others). These different processing strategies have to be treated with caution because they entirely rely on introspective data (Nisbet and Wilson, 1977). Nevertheless they offer an explanation for the unexpected result that all four resynthesised versions were equally acceptable in the non-contrastive contexts in the rating experiments.

4.5 Summary and Conclusion

In this chapter we described and discussed six rating experiments; Figure 4.6 summarises the course of the experiments, the main results, and important changes that became necessary are depicted.

First, we were interested in the relative perceptual importance of peak height and peak alignment, which differed significantly in the production data. These two variables seemed particularly interesting because subjects varied as to whether they made greater use of peak height or peak alignment. In order to investigate the influence of the respective variables we decided on resynthesised stimuli, which allow for good control with reasonable quality. However, the acoustic criteria for resynthesis were not immediately clear because of a strong interaction and unknown causal relationships between the various variables. So we decided to conduct the experiments with two different manipulations (shown in the upper panel of Figure 4.6).

For non-contrastive contexts, we initially chose the "start-of-a-story", which turned out to be too unrestricted. This problem was later addressed in the experiments displayed in the lower panel of Figure 4.6. For contrastive contexts, we chose a semantic parallelism where subjects had to judge the appropriateness of the intonation of the second clause with respect to the first clause. In contrastive contexts, the presumed non-contrastive version (with a low and early peak) was dispreferred over all other versions. It turned out that peak height (with the concomitant factor slope of the rise) was by far the most important contributing factor to the perception of contrastive themes. A later peak position, on the other hand, was not judged as appropriate as a higher peak. In these first experiments we solely manipulated the fundamental frequency. Since alignment differences are correlated with durational differences and since duration is a strong cue to prominence, we moved on to investigating the relative importance of peak height and segmental duration. We first presented a better-suited non-contrastive context, using a question-answer pair method-

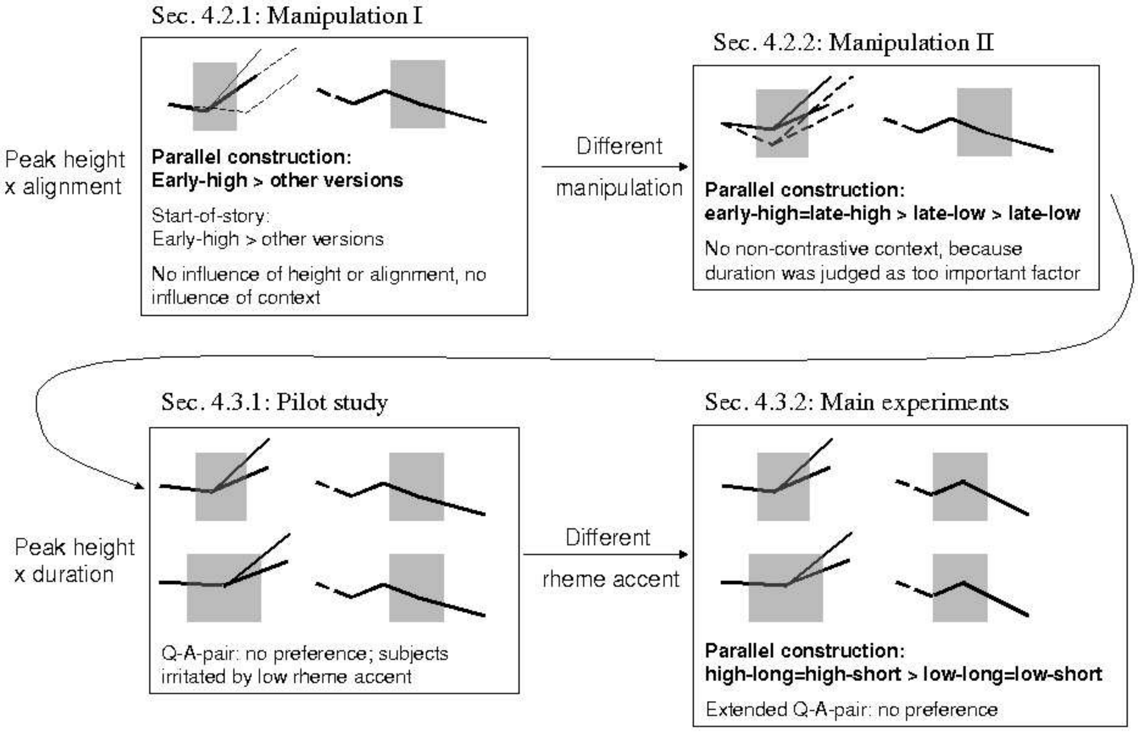


Figure 4.6: Overview of all rating experiments. Bold face indicates contrastive context, normal font non-contrastive one. Arrows show the advance of the experiments. Boxes are labelled with the section where a detailed description is given. In the experiments in the upper part, `peak height` x `peak alignment` was investigated, in the lower part, `peak height` x `duration`.

ology. This involved a change in rheme accent type, because a question-context favours a "newness"-accent. In these experiments, subjects again preferred versions with higher peaks in contrastive contexts, while they judged all versions as being equally appropriate in non-contrastive contexts.

Summarising the influence of the phonetic variables, a higher peak was a stronger cue to contrast than both peak alignment and segmental duration. Alignment differences only played a role in the low peak versions of manipulation II (*late-low* > *early-low*). As opposed to peak alignment, segmental duration could not compensate in the same way for the missing peak height to achieve a contrastive interpretation (*low-long* = *low-short*). However, the good discrimination between the intended non-contrastive and contrastive themes in contrastive contexts (which are all signalled by a rising prenuclear accent) might feed into the refinement of semantic theories. These, up to now, base their interpretation on rather general ToBI categories and largely ignore more detailed distinctions. While in contrastive contexts, little variation in theme-accent type was well perceived by listeners, there seemed to be a free variation in rheme-accent type (in contrastive contexts). Although the parallel constructions triggered both a contrastive theme and rheme, subjects accepted both high and low rheme accents if only the thematic accent was appropriately marked. This indicates that rhematic accents need not necessarily be marked by an (L+)H*-accent as commonly assumed (cf. Pierrehumbert and Hirschberg, 1990; Bartels and Kingston, 1994) but may equally well be marked by a low accent (!H+L*) in combination with a rising prenuclear accent. A similar observation is reported in t'Hart and Cohen (1973): "[...] we have been able to show that alternations of the contour do not inevitably and consistently result in changes of interpretation, as seems to be the generally adopted view" (p. 326). This provides further counter-evidence against a simple 1:1 mapping of tonal accents to semantic meaning.

The contrastive context allowed for freedom concerning the rheme accent marking but subjects had clear preferences for the thematic accent. On the other hand, the non-contrastive contexts demanded a high rheme accent but theme marking seemed largely irrelevant. This observed asymmetry between accent marking in these contexts might indicate that there are restrictions concerning the compositionality of accents. Restricted compositionality is not that uncommon in intonational structure: Syntactic constraints were also defined by Trager (1964) for the American Structuralist tradition. t'Hart et al. (1990) provided a "grammar" for specifying the combinatory possibilities of accents "that generates all and only the permissible sequences of pitch movements in a given domain, e.g. that of the clause." (p. 78). In addition, Grabe and Gussenhoven (1997) have shown for Dutch that there are distinct preferences concerning the combination of preheads with different pitch heights and nuclear accent types. Dainora (2002) proposed a probabilistic model of intonation in American English, based on ToBI categories (pitch and phrase accents and boundary tones). On the basis of her data she argues against a compositional approach for intonational meaning: "The data [...] constitute a strong rejection of any theory that suggests that tones are chosen independently of the tones that precede them [...] The strong interrelations among tones suggest that the meaning of a tune is more than the sum of its tones" (p. 238). Given these observations, there is a strong need to rethink present-day accounts for a compositional semantics of intonational events (Pierrehumbert and Hirschberg, 1990; Steedman, 2000).

While there were clear preferences in contrastive contexts, all versions were judged as equally appropriate in non-contrastive contexts. We evaluated two different hypotheses. First, semantically oriented, the presumed inappropriate versions could have been accepted together with an accommodation of an unexpressed contrast (Thomason, 1990). Second, along the line of Andreeva and Barry (1999), we assumed that non-contrastive contexts are prosodically under-specified and that intonational marking may be interpreted in terms of paralinguistic meaning apart from contrast or information structure. A forced choice experiment in non-contrastive contexts (in which subjects additionally had to explain their choice) showed that in most of the cases subjects preferred the presumed non-contrastive versions. Explanations given were mainly information structural. In the rarer cases, in which subjects chose the presumed inappropriate (contrastive) version, they used explanations that mostly made inferences about positive "speaker" attitudes, such as friendliness, enthusiasm, and interest.

The different results obtained in contrastive and non-contrastive contexts do not allow us to claim a direct relation between intonational marking and contrast. Also, analysing the perceptual importance of the different acoustic variables (such as peak height, peak alignment, and duration) does not tell us whether the acoustic differences observed in contrastive and non-contrastive contexts can be annotated using different accent types. A reliable prosodic annotation would be a prerequisite for present-day semantic formalisms. In the next chapter we shall present data from a labelling experiment and discuss the implications of our data for the prosody–semantics interface.

Chapter 5

Impacts for the Prosody–Semantics Interface

To use prosodic information in semantic formalisms we need to define an interface between these two layers. These interfaces generally use some sort of *categories*, such as phonemes or morphemes. In this chapter we shall discuss whether the prosodic marking of contrastive and non-contrastive themes can be described with different accent types or not. The results of the production experiment showed that there are significant acoustic differences between themes in contrastive and non-contrastive contexts. However, it has been noted by several linguists (Bolinger, Gussenhoven, Ladd, among others) that one main difference between prosodic and segmental events is that the former are not based in the same way on categorical contrasts as the latter are. This non-categorical nature of prosodic events can be either explained as free variation (possibly influenced by contextual factors) or as conveying semantic and/or pragmatic meaning. Ladd (1980), for instance, pursued the view that prosodic variation — or in his terms *gradience* — in intonational form is supposed to reflect gradience in semantic (or pragmatic) meaning:

> "The noteworthy feature of intonational meaning that has always frustrated the efforts of segment-minded linguists is the fact that much of it seems to be scalar, or more precisely, that intonation matches semantic continua with formal ones." (Ladd, 1980, p. 106)

As Ladd further noted, we have to establish if the semantic phenomenon we are interested in — thematic contrast in our case — exhibits distinct intonational categories.

> "[...] the principal taxonomic task of the investigator is to determine which of the semantic distinctions of intonation represent contrast between linguistic categories and which represent gradient variation within categories." (ibid, p. 112)

From a purely semantic perspective, thematic contrast is a linguistic phenomenon with a categorical (all-or-none) nature: either there is a contrast to a previously mentioned alternative theme or there is none. However, thematic contrast may alternatively be interpreted as topic resumption or enhanced emphasis, which are neither entirely non-contrastive nor contrastive but based on fine-grained discourse-structural distinctions. Cognitively, contrast-marking functions successfully as a highlighting operation for alternatives (section 2.3.1), which, however, need not be categorical; we might equally well assume different *degrees* of highlighting (which could be related to the cost of retrieving the respective items).

In this chapter we first address the prosodic side of the prosody-semantics interface (section 5.1), for which we already gathered evidence in both production and perception experiments. We first review some experimental results gained in the classical categorical perception paradigm (section 5.1.1) before we present an experiment of our own in which we mainly investigated whether there is a categorical difference in thematic contrast marking that can be used in semantic formalisms. We asked phoneticians that are experienced in using GToBI as a tool for prosodic annotation to label the prenuclear accents of a subset of our data (section 5.1.2). In section 5.2 we then briefly outline the implications of the present results for semantic theories.

5.1 Prosodic Categories

The results of the perception experiments with linguistically naïve listeners reported so far do not offer conclusive evidence for or against a categorical distinction. In chapter 4, we used rating studies to test the perceptual importance of the variables that differed significantly in the production experiment (peak height, peak alignment, and segmental duration). In contrastive contexts, subjects could reliably use intonational information in their judgements. However, the results of the appropriateness ratings showed that listeners were especially susceptible to pitch height differences, a variable that is generally said to be gradually variable. This is problematic for the prosody–semantics interface because it rather relies on categorical differences. In this section we shall discuss ways to evaluate whether there are distinct categories involved.

We first discuss a specific perception paradigm to investigate the categorical nature of events, which bears the name of the phenomenon it is assumed to show: categorical perception (section 5.1.1). We shall describe what is exactly meant by categorical perception and discuss the method behind it. Although categorical perception was first established for consonants (cf. Liberman et al., 1957), it has since been used to study the categorical nature of intonational events (such as peak height and peak alignment). However, there is still some dispute about its applicability, especially, because some intonational phenomena, such as emphasis, seem to be inherently non-categorical as Ladd (1980) pointed out. Recently, Schouten et al. (2003) have criticised the lack of operational objective criteria for *interpreting* the results of categorical perception experiments.

We therefore decided to pursue an alternative approach that additionally allowed us to evaluate the phonetic basis for building up an information-structural semantic analysis (section 5.1.2). We investigated GToBI-labellers' abilities to annotate the prenuclear accents of a subset of our data. Since these accent types are categorically distinct *per definitionem*, we should be able to infer from the labels annotated whether themes in contrastive and non-contrastive contexts are categorically distinct (judging from the number of labellers that assign the same or different accent types to the two versions of an utterance pair). Furthermore, if different annotators were able to label a major part of the data consistently, we could be confident in using information about pitch accent types for semantic theories. If, on the other hand, labelling turned out to be rather inconsistent, we should rethink either the applicability and granularity of the annotation scheme used or the validity of current information structural semantics.

5.1.1 Review: Categorical Perception Experiments

The phenomenon of categorical perception (CP) was first postulated in the fifties in psychophonetic research for the perception of plosive phonemes (cf. Liberman et al., 1957). Broadly speaking, categorical perception covers the fact that manipulated tokens that belong to the same category are judged as being the same even if they differ acoustically. Furthermore, subjects should not be able to perceive the differences between two different tokens of the same category (i.e. the ability to discriminate between phonemes should be completely determined by the ability to categorise them). Although not demonstrated in this absolute form, this special sort of perception was then taken as evidence for a special status of speech compared to non-speech. Since the 1950's, approximations to categorical perception have been established for a variety of perception areas, including vision and even social perception (Harnad, 1987).

In the original experiments on categorical perception, it was shown that subjects did not perceive gradual changes in a continuously changing variable (such as voice onset time, VOT). However, their perception changed abruptly when VOT exceeded a certain limit. Furthermore, they had much enhanced discrimination for stimuli whose VOT lay on either side of this limit. These are generally the premises to postulate that subjects are able to categorise the continuously varying speech sounds (here into e.g. voiced or voiceless plosives). Normally, a series of more than ten stimuli is created where the variable under investigation is changed in equally-sized steps to create a pseudo-continuum[1]. The experiments involve two distinct perception tasks, a classification and a discrimination test. In the classification task, which is also referred to as identification task, stimuli have to be assigned by forced

[1]Note that the manipulations are one-dimensional, such as closure duration.

choice to one of two categories given. In the discrimination task, two adjacent stimuli (in some cases stimuli that are two steps apart in the pseudo-continuum) have to be rated as being the same or different[2]. If a given variable is indeed perceived categorically (as in the case with increasing VOT and the change in perception from /b/ to /p/), the classification results for one of the categories (e.g. the perception of /p/) should result in an S-shaped contour, indicating an abrupt change in perception (from the category /b/ to /p/). In contrast, the perception of /b/ abruptly decreases at the cross-over point. In addition to the identification characteristics, results of the discrimination task should result in a discrimination peak at the area where the change in perceived category takes place. That is, at the category boundary, discrimination between stimuli is enhanced, but within a category, discrimination is hardly possible.

There have been attempts to use this paradigm for intonational phenomena in almost all frameworks of intonational phonology, American structuralists, British School, and autosegmental-metrical phonology. The most promising distinction for finding categorical perception for intonational events is the interpretation of boundary tones (statements vs. questions) as investigated by Hadding-Koch and Studdert-Kennedy (1964); Remijsen and van Heuven (1999), among others[3]. Also for certain pitch accent distinctions, there are results in the literature that suggest a categorical perception. Kohler (1991b), for instance, investigated the categorical nature (and the corresponding semantico-pragmatic effects) of early, medial and late peaks in German (in autosegmental terms, these accents might be described as !H+L*, H* and L+H*). They resynthesised eleven peak positions in the single-accent utterance “Sie hat ja gelogen.” (*“She has been lying.”* or *“She lied, of course”*). There was a categorical distinction between early and medial peaks (pragmatically indicating an “established fact” as opposed to “new information”) but there was no categorical distinction between medial and late peaks, although late peaks were perceived as more emphatic (for further details see Kohler, 1991b).

Ladd and Morton (1997) tested the categorical perception of different peak heights along the pragmatic dichotomy “everyday occurrence” and “unusual experience” (p. 325). They managed to replicate an S-shaped identification contour but failed to reproduce the discrimination peak. That is, subjects are discriminating different meanings but they also discriminate different realisations *within* a “category”. Ladd and Morton’s conclusion was that “[w]hile we cannot therefore legitimately speak of pitch range being categorically *perceived*, it may be useful to think of them as being categorically *interpreted*.” (Ladd and Morton, 1997, p. 339). That is, compared to segmental perception, the perception of in-

[2]Especially for the discrimination task, there are various methods which are not discussed here further. The interested reader should check Repp (1984) for a concise summary.

[3]Note that Hadding-Koch and Studdert-Kennedy (1964) did not perform a discrimination task; they report only an identification task and a psychophysical judgement where subjects had to indicate whether the utterances ended in a high or low pitch.

tonational events is probably more susceptible to fine variations in the acoustic structure. One might therefore argue that intonational meaning is not as robustly defined, produced and perceived as segmental meaning distinctions. This is easily be explained by the multi-functionality of prosodic structure. If various semantic dimensions (e.g. utterance mood, attitude, dialogue act, emotion) have to be encoded on the same channel, the single "categories" cannot be defined that strictly. That is, subjects might perceive a difference within categories because it is simply a difference in another (possibly paralinguistic) function.

Remijsen and van Heuven (1999) explained Ladd and Morton's lacking discrimination peak by classifying it as paralinguistic. Others propose a continuum between linguistics and paralinguistics (e.g. Crystal, Bolinger, Ladd, Gussenhoven). If there is an all-or-none contrast, then it is clearly linguistic and we should find a categorical perception. If, on the other hand, a gradience in form implies a gradience in meaning (e.g. more or less emphatic, more or less uncertain, etc.), then subjects should be able to perceive the subte differences between the versions presented.

The discussion about the continuous or categorical nature of peak height distinctions has received further impulses from a recent study by Chen (2003). She replicated the experiment of Ladd and Morton (1997), but instead of a discrimination task she simply measured the response time in the identification task (from the start of the stimulus presentation to its identification). Her reasoning was that identification should be faster within categories than at the boundary between two categories (cf. Pisoni and Tash, 1974). And indeed, the course of reaction time is fairly low within categories and only increases near the category-crossing. This function resembles the discrimination function of categorical perception experiments, except for a short plateau instead of a "peak" in discrimination. This can be interpreted as some support for a categorical distinction of peak height[4].

This short overview has shown that it is not completely clear how to *assess* the perception of intonational events and how to *interpret* the results. Schouten et al. (2003) further criticised that there are no explicit criteria for interpreting the results of categorical perception experiments: "anything resembling a discrimination peak anywhere near a phoneme boundary has generally been enough to claim categorical perception of the stimuli concerned" (Schouten et al., 2003, p. 72). They investigated several factors that might influence the results of categorical perception experiments: naturalness of stimuli, inter-stimulus interval, and response bias. They concluded that "[t]he most crucial factor, however, is the discrimination task. A task that compels listeners to refer the stimuli to a subjective criterion is bound to produce results that indicate some degree of categorical perception; a task that is free of bias will produce no categorical perception at all." (ibid, p. 79). Kohler

[4] In a parallel experiment, she investigated the perception of different peak alignments in the utterance *'To Birmingham?'*. Subjects had to decide whether it was a neutral request for information or an incredulous echo-question. In contrast to peak height differences, subjects were not able to identify the utterances consistently, nor was there an influence on reaction times.

(p.c.) claims that the discrimination peak at category crossing merely is a special case in perception. Because of this ongoing debate about the applicability of categorical perception experiments (especially for intonational phenomena), we decided against this method. As discussed in section 4.1, a further problem would have been to understandably define the "category" of contrast and non-contrast for the identification task.

5.1.2 GToBI Labelling Experiment

The GToBI labelling experiment reported here had two closely related purposes. First, we wanted to find out whether GToBI-labellers could make a categorical distinction between prenuclear rises that were produced in contrastive and non-contrastive contexts. Therefore we asked them to label a subset of the utterance pairs recorded using GToBI (German Tone and Break Indices, see section 2.1.1 for an overview). Since autosegmental-metrical accent types are categorically distinct, labellers should consistently choose a *different* accent type if there was a categorical distinction in the data. This experiment should help us to verify the prosodic basis for current semantic formalisms (e.g. Büring, 1997; Steedman, 2000), which rely entirely on specific accent type distinctions to build up an information structural semantics.

In previous experiments we had mostly used resynthesised stimuli to maintain good control over the prosodic variables. This procedure relied on two assumptions; first that naïve listeners do not possess analytic perception abilities to detect the variables varied and second that we know exactly which variables to manipulate. The first assumption is not valid for trained phoneticians. Furthermore, natural speech stimuli — with a strong interaction of a variety of prosodic variables — are exactly the material semantic formalisms should ideally be able to cope with.

Goal. The labelling experiment described in this section was exploratory: We aimed at finding evidence for or against the view that the observed differences in theme marking resulted in discrete tonal categories. We reasoned that if many annotators labelled the two realisations of a given utterance pair with the same accent type, this would indicate that the differences we had found in the production data constituted mere phonetic variations within one accent type, and this cannot be expressed with standard autosegmental categories. If, on the other hand, annotators consistently labelled the two members of a given pair with different accent types, this would provide evidence for a categorical distinction. In that latter case, we could explore further on which phonetic variables this change in percept is based. Irrespective of the phonetic properties, semantic formalisms could safely employ this categoriality.

Stimuli. From the 86 utterance pairs, we randomly selected ten pairs from different speakers that differed in peak height and peak position, see Figure 5.1.

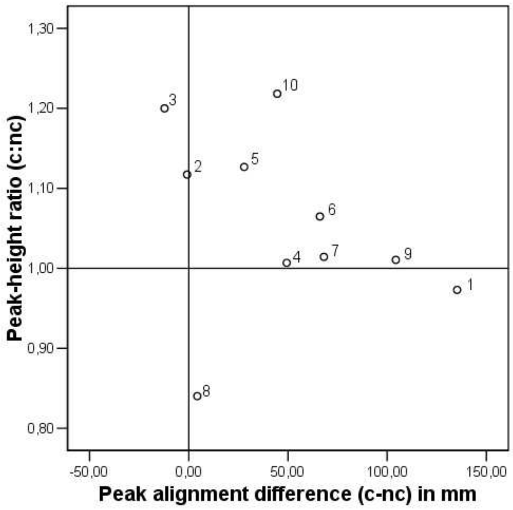


Figure 5.1: Peak-height ratio and peak alignment difference of the ten sentence pairs used in the labelling experiment.

There was an equal number of preposition and noun phrases in the preverbal field (see Table 5.1). Although the rheme-accent of a given utterance should not influence the labelling of the prenuclear accent, we only chose pairs in which the position and form of the rheme accent was identical in the contrastive and the non-contrastive realisation. In most utterance pairs, the rheme-accent was marked by a H+L*, but our sample also contained utterance pairs with a high nuclear accent[5]. We decided on this preselection of material because we wanted to keep the test small and concise to attract as many labellers as possible.

Method. 12 annotators from different German universities, who were experienced in using GToBI for prosodic annotation, were individually contacted via email and asked to participate in this anonymous, web-based labelling experiment. After starting the experiment, the prenuclear accents of the 20 stimuli had to be labelled in random order. Labellers were not explicitly pointed to the presence of utterance *pairs* but they nevertheless might have detected them[6]. We decided against a pairwise presentation because subjects could have been inclined to mark the differences they perceived even if these differences would not exceed the threshold of categories, if presented in isolation[7].

[5]We do not want to enter into a discussion about downstep. We generally indicated the nuclear accent with H* but it may be that some of the accents are actually !H* accents.

[6]In order to disguise the presence of utterance pairs we would have had to present much more stimuli, which was avoided for the sake of a short experiment.

[7]This was for example observed in the discrimination task in the study of Ladd and Morton (1997): Subjects were able to categorise stimuli with different pitch heights to either “everyday occurrence” and

Stimulus-pair	gramm. rel.	rheme accent
BK_armenien	PP	H+L*
BK_milano	PP	H+L*
CH_laemmer	NP	H+L*
CS_bayern	PP	H+L*
CS_milano	PP	H*
DS_januar	PP	H+L*
DS_roemer	NP	H*
IB_italiener	NP	H+L*
JL_laemmer	NP	H+L*
UB_italiener	NP	H*

Table 5.1: Utterance pairs used in the GToBI labelling experiment. The 2nd column shows the grammatical relation of the element in the preverbal position. The rheme-accent, which was the same for both realisations of a pair, is shown in the last column.

Subjects were told that the experiment was meant to collect intersubjective labels of 20 prenuclear accents. 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[8]. It was described how to download the files for visual inspection of the f_0-trace. For labelling, 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" that they could specify in a special text field nearby. Furthermore, we asked them to rate the certainty of their judgement on a scale from 1 (very uncertain) to 5 (very certain) and specify the reasons that contributed most to their decision. In addition, we asked them to indicate whether they had used visual information about the f_0-contour.

Results and Discussion. Of the twelve labellers contacted, eight completed the experiment. 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. If we count only those items that received the least 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). Since only few labellers indicated the reason for their uncertainty (e.g. by naming alternative accent types

"unusual experience", but could still perceive a difference between two neighbouring stimuli (with slightly different peak heights).

[8]The judgement of isolated utterances is harder than labelling longer stretches where "labellers have access to enough context for 'tuning in' to each speaker's range and vocal characteristics" (Grice et al., 1996, p. 1717).

that they assumed to be also probable), the certainty values are not considered further in the analysis.

Only one of the labellers indicated that she had used visual information about the f_0-movement (more specifically about the alignment between the f_0-peak and maximum in energy). One labeller stated that she had imagined the course of f_0-movement. Another labeller reported that he took semantic considerations into account. In one case, a phonetically plausible accent type was rejected for semantic reasons[9]: “Aus semantischen Gründen nicht L*+H. Bei rein phonetischer Verwendung von ToBI wäre auch L*+H erwägenswert.” Finally, one labeller (labeller 7) was very accurate and often used diacritics, especially for indicating if an accent was late, which we preserved in the analysis. Late accents were also perceived by other labellers, but often not annotated (“tiefer Beginn auf ‘die’, hoher Zielton wird allerdings erst spät auf der zweiten Silbe von ‘Lämmer’ erreicht, daher etwas unsicher in der Entscheidung”); they were obviously not seen as distinctive by most of the annotators (“Trotz spätem Gipfel L+H*”).

On average, in 52.5% of the cases, the two versions of an utterance pair were annotated using the same accent type, whereas in the remaining cases, the two versions of a pair were labelled with different accent types.

Figure 5.2 shows the accent types labelled for the ten sentence pairs. The size of the bars in the *upper part* of the figure corresponds to the number of annotators who labelled the two versions of an utterance pair with the *same* accent type. It becomes clear that “same accent” is not equivalent to “intersubjectively identical”; rather, different experts used different intonational categories to express their percept. Even in the cases in which six out of eight subjects labelled both versions of an utterance pair with the same accent type (pairs 1, 2, and 8), there were different opinions on the type of accent to assign, with a pronounced preference for L+H* (22 $\times$) as opposed to L*+H (15 $\times$), H* (4 $\times$), and H*+$\hat{}$H (1 $\times$).

The lower part of Figure 5.2 shows the labels for the cases in which the annotators labelled the two versions of a sentence with different accent types. There was one very common distinction that occurred 17 times: L*+H (contrastive) vs. L+H* (non-contrastive). Less preferred but still very frequent was a label opposition between L*+H vs. H* (8 $\times$). Here also, the agreement between annotators was rather low.

The range of different accent types used indicates 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. In other words, it seems quite likely that there actually exists a perceptual differ-

[9]Note that we counted the phonetically appropriate L*+H in that case.

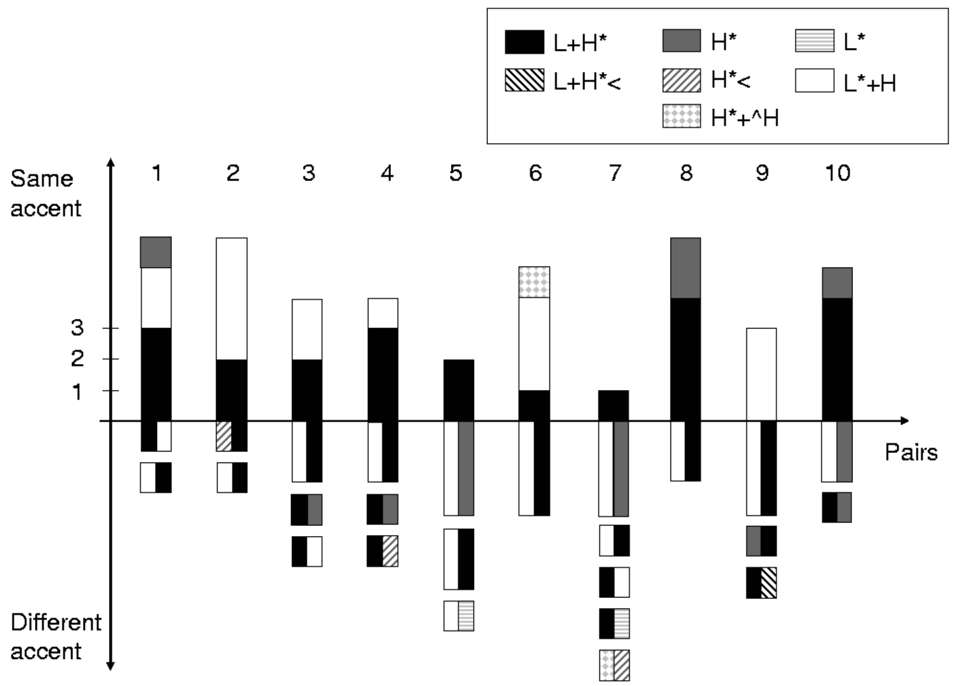


Figure 5.2: Graphical display of the results of the labelling experiment. The ten utterance pairs (identical utterances produced in contrastive and neutral contexts) are plotted along the x-axis. The height of the bars in the upper part of the chart corresponds to the number of annotators that labelled both versions of the pair with the same accent type. The height of the bars in the lower part of the figure represents how often the two versions of the pair were annotated with *different* labels (left bar represents the realisation in contrastive context, right bar the realisation in non-contrastive context). Different "label-pairs" are separated vertically by small spaces.

ence between the contrastive and non-contrastive realisation of these sentences. For the moment we ignore the diversity in the accent types used and concentrate on the number of annotators that used the same or different accent types for the two versions of a pair. To find out to which acoustic differences labellers were most sensitive, we correlated the number of labellers who annotated the two versions of an utterance pair with a different accent type with variables that describe the acoustic difference between the realisation in contrastive and neutral contexts (such as difference in peak alignment between contrastive and neutral realisation, peak height-ratio and peak excursion-ratio between contrastive and neutral realisation and slope-ratio of the rise); the values are shown in Table 5.2.

No.	Stimulus-pair	number of "same accent"	peak-height-ratio	slope-ratio	f_0-excursion-ratio	al-diff (H^*,V1) in ms
1	BK_armenien	6	1.06	1.31	1.03	21.5
2	BK_milano	6	1.11	1.27	1.30	- 16.6
3	CH_laemmer	4	1.20	1.48	1.28	- 24.9
4	CS_bayern	4	1.56	1.35	1.43	49.3
5	CS_milano	2	1.13	1.70	2.49	27.9
6	DS_januar	5	1.16	1.15	1.91	12.4
7	DS_roemer	1	1.01	1.04	2.28	67.5
8	IB_italiener	5	1.13	1.35	1.41	19.1
9	JL_laemmer	3	1.01	0.82	1.39	104.4
10	UB_italiener	5	1.21	1.05	1.17	100.8

Table 5.2: Utterance pairs presented in the labelling experiment, together with the number of annotators who assigned the same accent type to both items of the pair. In the last four columns, the peak height ratio between contrastive and non-contrastive version, the slope ratio, and the f_0-excursion-ratio as well as the difference in peak alignment between contrastive and non-contrastive realisation are presented.

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 little 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 exists with peak alignment differences, the genuine basis for intonational contrasts in AM-approaches since Bruce (1977). It would have been natural to expect that the larger the difference in alignment between the two versions, the more labellers would label the two versions with a different accent type. GToBI offers diacritics to indicate that the pitch range is expanded (upstep ^H) or decreased (downstep !H). These diacritics only describe changes in pitch range relative to an earlier high target and can hence not be used for sentence-initial accents. Therefore, it might be that the high degree of disagreement in accent types results from the fact that standard GToBI categories for prenuclear accents do not take the scaling of accents into account. 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 characterisitcs of the speakers.

Labellers might have their own internal "representation" of the ToBI-categories used. These 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 are indeed employed on a subjective, idiosyncratic basis, individual labellers should maintain a reliable difference between the contrastive and non-contrastive versions (if there is such a difference). We therefore analysed the results of the labelling experiments for the individual annotators separately. The result is shown in Figure 5.3.

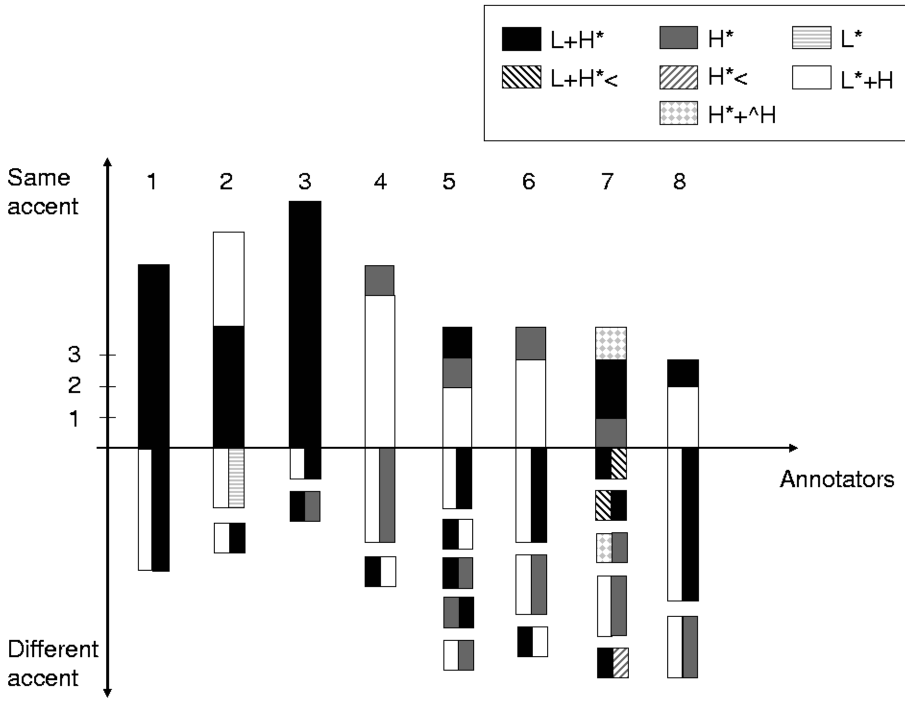


Figure 5.3: Annotator-specific results of the labelling experiment. When bars are in parallel, the left one shows the label annotated for the contrastive version.

There is a clear split in labelling behaviour. The first four labellers, for example, annotated the two versions of the majority of the utterance pairs with the same accent type, whereas the last four labellers used more different accent types. Most labellers are rather consistent in using L*+H in contrastive contexts (labellers 1, 2, 3, 6, and 8) but they used different accent types for non-contrastive contexts. Also, there appear to be idiosyncratically preferred prenuclear accent types (or biases). For some labellers, this is L*+H, for others it is L+H*. Labeller's choice of a particular accent type might be influenced more by the demands of the system they adhered to than by the auditory properties of the stimulus.

To conclude the results of the labelling experiment, it seems that intonational contrasts used in semantic formalisms, such as L*+H vs. L+H*, are indeed often employed by annotators to discriminate the realisations in contrastive and non-contrastive contexts. However, the annotation is not very reliable. Given the high amount 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 intonational form to thematic contrast seems unlikely.

Automatic Transcription. Since automatic transcription might be less biased than human annotation, we briefly discuss the results of an automatic transcription of our data. Braunschweiler (2003) implemented a system called ProsAlign that uses only information about the f_0-course, the intensity (root-mean-square, RMS) and voicing information to assign ToBI-style pitch accents and boundary tones. Unlike other automatic annotation systems (cf. Wightman and Ostendorf, 1994), Braunschweiler does not use statistical techniques such as Hidden-Markov Modelling (HMM) but pursues a rule-based approach. Thanks to Norbert Braunschweiler, we had access to the automatic transcription of the 20 stimuli that were labelled by our annotators. The automatic transcription of the accents are presented in Table 5.3.

Stimulus-pair	non-contrastive context	contrastive context
BK_armenien	**L+H***	
BK_milano	L*	L+H*
CH_laemmer	no acc.	L*+H
CS_bayern	**H***	
CS_milano	**L+H***	
DS_januar	no acc.	H*
DS_roemer	H*	L+H*
IB_italiener	**H***	
JL_laemmer	L*	L+H*
UB_italiener	H*	L+H*

Table 5.3: Automatic transcription of the prenuclear accents by ProsAlign (Braunschweiler, 2003) for the ten utterance pairs. The label for the non-contrastive realisation is shown to the left, the label for the contrastive version to the right. Labels in bold-face represent the cases for which ProsAlign chose the same label for both realisations.

ProsAlign sometimes indicated two accents in the preverbal position, so we had to chose the correct one after inspecting their positioning with respect to the segmental structure.

For two of the non-contrastive stimuli, ProsAlign did not find any accent at all (item 5, CH_laemmer and item 11, DS_januar). This might have been caused by insufficient energy in that area. Compared to human labellers, ProsAlign more often assigned different accent types to the two versions of an utterance pair (if we count the missing accent as "no accent"). Table 5.3 lists the labels assigned by ProsAlign. If there is only one label in a row, the respective accent type is assigned to both realisations. Comparing the labelling behaviour of the transcribers and ProsAlign for all ten utterance pairs shows both agreement and disagreement. For pairs 1 and 8, most of the annotators labelled the two realisations with the same accent and so did ProsAlign. On the other hand, ProsAlign also annotated the same accent type for the two versions of pair 5, which were labelled with different accent types by the majority of human annotators.

5.2 Discussion

Grice and Baumann (2002) claim that GToBI is a transcription system that is easy to acquire and flexibly usable (p. 275). However, an earlier conducted consistency experiment with 13 annotators (Grice et al., 1996) showed considerable difficulties in distinguishing certain accent types. Although the labellers were trained on the task by special training material (cf. Benzmüller and Grice, 1997, now available under http://www.coli.uni-sb.de/phonetik/projects/Tobi/gtobi.php3), annotators' agreement on pitch accents (incl. zero accent) was 71%, which was slightly better than the results reported in Pitrelli et al. (1994) for English. When transcribers agreed on the presence of an accent, agreement on the type of accent was 51%. The accent pairs that resulted in highest disagreement were L+H* vs. H* (33%) and L+H* vs. L*+H (14%). That is, the annotation of rising accents seems to be particularly difficult and hence untrustworthy, which is confirmed by our transcriptions. Our analyses are clearly preliminary and hence do not allow strong claims, because they are based on data from only eight annotators. Nevertheless, it has raised some doubt about the definition of the GToBI accents, and, more importantly, about their alleged semantic function.

It turned out that the number of transcribers that assigned a different label to the two versions of an utterance pair increased with an increasing difference in f_0-excursion between the two versions[10]. There is hence some indication for a perceptual importance of f_0-excursion on the assignment of tonal categories. However, f_0-excursion — or more generally pitch range — is not a criterion for distinguishing between tonal categories in GToBI. In the latest description of GToBI[11], Grice and Baumann (2002) list the following

[10]Note that we shall not examine further why particular annotators chose certain accent types. Rather, we are interested in the cases where the majority of annotators labelled the two versions of an utterance pair with the same or with different accent types.

[11]For an English translation see Grice et al. (2005).

criteria for the accents:

H*: accented syllable perceived as high, preceded by a flat rise.

L+H*: accented syllable perceived as high, which immediately follows a low target. There is a steep rise in f_0 or a jump up to the accented syllable. The peak is often reached late in the accented syllable.

L*+H: a low target within the accented syllable is followed by a (steep)[12] rise that starts late in the accented syllable and reaches its peak on the following syllable (sometimes later). In contrast to L+H*, the accent is perceived as low.

There is no mention of pitch excursion in the definition of the accent types. However, the difference in f_0-excursion between the two versions of a pair was the factor that correlated most with the annotation of different or same accent labels to the two versions, and some labellers reported especially high pitch excursions (e.g. "Stimme überschlägt sich fast, so hoch geht der Sprung").

In nuclear position, there are more possibilities to describe a rising movement by making use of phrase accents and boundary tones. This also allows for describing different pitch heights, e.g. L* L-H% vs. L* H-^H% where the latter contour has a higher end point than the former one. ToBI further supports the possibility of annotating differences in the scaling of accents: upstep (signalled by ^H) and downstep (signalled by !H), but these diacritics are only allowed *after* a high tone. That is, upstep and downstep are not available for the phrase-initial pitch accents we are dealing with.

Pitch range phenomena appear in one way or another in all linguistic transcription systems, in the American structuralist approach to intonation description (e.g. Pike, 1945; Trager and Smith, 1957; Moulton, 1962; Isačenko and Schädlich, 1966) as well as in perceptually oriented intonational models, such as the British School (Kingdon, 1958; O'Connor and Arnold, 1961; Crystal, 1969; O'Connor and Arnold, 1961; von Essen, 1964; Pheby, 1975) or the Dutch model (t'Hart et al., 1990)[13]. In the American structuralist's tradition, peak height is described in terms of discrete levels. That is, rises and falls are decomposed as pitch movements from one level to another. While this resembles modern autosegmental-metrical approaches, the American structuralists used four different pitch levels, defined on a relative basis[14]. This allows for describing f_0-movements with varying pitch excursions (e.g. /1-4°/ vs. /2-4°/ or /1-3°/). In the tradition of the British School as well as in the Dutch school of intonation (t'Hart et al., 1990), configurations with the same movement (fall, rise, fall-rise) are distinguished with respect to the f_0-excursion. O'Connor and Arnold

[12] Round brackets also present in the original (Grice and Baumann, 2002, p. 278).

[13] These models were described in more detail in section 2.1.1.

[14] There are more striking differences between the level-approaches from the sixties and seventies and autosegmental-metrical approaches, which were discussed in more detail in section 2.1.1.

(1961), for instance, state two falling and two rising contours: *Low Fall* vs. *High Fall* (the former starting from a medium pitch, the latter from a high pitch) and *Low Rise* vs. *High Rise* (starting from a low or medium pitch). Especially this latter distinction between *Low Rise* and *High Rise* seems to be appropriate to reflect the criteria our transcribers probably based their decisions on: While peak *height* differences seemed to play no role, peak *excursion* (probably achieved by a lower starting point) influenced the choice of category.

This raises the question whether we really *have* to account for accents with different pitch excursions and pitch ranges in intonational phonology for sentence-initial accents, and hence in annotation systems. The inconsistency observed in the present GToBI-annotation suggests that the present system of the annotation is all but easy to use. It is questionable, however, whether integrating information about the scaling of (initial) accents into an autosegmental-metrical approach would make labelling more consistent. Instead, Ladd has repeatedly made the case for gradient differences in form that might reflect gradiently variable meaning differences (Ladd, 1980, 1996). And he warned against the danger of deliberately not accounting for pitch range phenomena. And our data show that pitch range excursion is used in a linguistically meaningful way.

> "[...] because variation in the vertical scaling of pitch accents *may be* a matter of unpredictable, paralinguistic, gradient variation, phonologists generally assume that any variation in vertical scale *is* a matter of unpredictable, paralinguistic, gradient variation, and consequently do not look more closely at cases where the vertical scaling may actually be subject to more systematic constraints." (Ladd, 1994, p. 44)

What are the implications of our empirical findings for semantic theories? The analysis of the production data has shown that the hat pattern is not found more often in contrastive contexts compared to non-contrastive ones (which runs counter the prosodic foundations of the semantics proposed by Büring, 1997). That is, semantic formalisms have to be more fine-grained to account for thematic contrast. Steedman (2000) proposed a compositional semantics for English, using autosegmental-metrical categories to account for thematic and rhematic contrast (broadly speaking, L+H* LH% for contrastive themes and H* LL% for contrastive rhemes). Autosegmental-metrical annotation schemes easily qualify for building a concise compositional semantics for intonational meaning (Pierrehumbert and Hirschberg, 1990; Hobbs, 1990; Steedman, 2000), but a valid compositional semantics of pitch accents presupposes that annotators are able to consistently assign the different accent types used in the semantic descriptions. The GToBI experiment reported in this section revealed considerable inconsistency among trained labellers, which undermines the validity of semantic claims.

We do not have conclusive evidence for or against a categorical difference in theme marking but there are two possible ways of future research: First, the read utterance pairs are probably no prototypical realisations of contrastive and non-contrastive themes and further — more natural — productions are necessary. In this respect it is a promising finding to have three utterance pairs (5, 7, and 9), for which the majority of labellers annotated the two versions with a different accent type. This might indicate a categorical distinction[15], which cannot be reliably annotated with present-day annotation schemes, however. It appears that the prosodic categories are not yet defined well enough to be straightforwardly used in semantic formalisms and much more research is necessary to investigate which phonetic variables exactly influence the differences in perception. Only then can we think about refining prosodic annotation schemes to account for the perceived differences, e.g. by annotating a difference in scaling of initial accents. If these diacritics could be reliably annotated, semantic formalisms merely needed to alter the prosodic categories they refer to. Second, it might be wrong to eagerly search for a categorical distinction between contrast and non-contrast because we could rather be dealing with a continuum (or pseudo-continuum) in discourse-functions, such as topic-continuation, topic-resumption, topic-change, etc. It is difficult to pursue further research in this direction in the near future because neither the discourse-structure side nor the prosodic side is theoretically well equipped to deliver independent results (often, e.g., intonation is used as a criterion for assigning different discourse-structures). Therefore, we might rather relate different degrees of prenuclear marking to "different degrees of activation". According to Chafe (1994), different costs are necessary to activate the discourse elements. Baumann and Grice (2004), in this vein, investigated the appropriateness of different nuclear accent types for the marking of different kinds of textually and inferentially given referential expressions and concluded that "the findings point to a scale or continuum of intonational marking, along which differing degrees of activation are expressed" (p. 24). In terms of *thematic* material, the costs for activating a new topic is likely to be higher than for a contrast with a related topic. Alternatively, an inferable topic can probably be activated with less cost than an entirely new topic.

[15] The present results do certainly not indicate a categorical perception in the strict sense of the term but probably rather some sort of categorical interpretation (in the sense of Ladd and Morton, 1997). This is further supported by the rating experiments with linguistically naïve listeners described in chapter 4. They showed some sensitivity for differences in intonational meaning as they consistently preferred stronger marked themes in contrastive contexts over less intonational marking.

Chapter 6

Discussion

In this doctoral dissertation we first devised a working definition for contrast. In the past, contrast has been defined by rather vague criteria, such as exhaustiveness or alternative sets. Those criteria led to the unsatisfactory situation that contrast is associated with information structural focus on the one hand and with corrections on the other hand. In our view, however, both of these associations miss the point. Contrast is clearly distinct from information structural focus, as especially examples from Hungarian showed. Furthermore, contrast is not restricted to rhematic, new material but can be established between thematic elements as well. This further undermines the commonly assumed equation between contrast and focus. We argued that rhematic contrast without an accompanying thematic contrast is hardly distinguishable from information structural focus (neither theoretically nor prosodically), which has probably contributed most to the reported confusion. From the literature on the phonetic marking of contrast it seems that this problem was intuitively clear to scientists; there, contrast is almost entirely elicited via corrections. We showed, however, that corrections are also quite different from contrast, as they strongly interfere with the grammatical level in some languages (e.g. Italian or Catalan) or are realised in a very hyperarticulated way (e.g. in English or German). Only in contrastive theme–rheme pairs (which we also refer to as *semantic parallelism*) is contrast sufficiently defined to draw a distinction to information structural categories on the one hand and to corrections on the other hand. Given our working definition of contrast, the focus of this thesis lay on the realisation and perception of *thematic* contrast. In addition, we analysed the behaviour of rheme accents, because to date there is only impressionistic data regarding the realisation of contrastive rheme accents in a semantic parallelism.

In this chapter we first summarise the main insights of the experimental studies on the production and perception of thematic contrast (section 6.1), before discussing selected theoretical aspects of the present work in more detail (section 6.2). There, we also consider how the interpretation of contrast might be assessed more directly using eye-tracking data gathered in the visual world paradigm.

6.1 Summary of Experimental Results

In the production experiment described in chapter 3, eleven subjects were asked to read aloud 54 paragraphs. Some of these paragraphs contained identical utterances, which were presented in both a contrastive and a non-contrastive context. More specifically, the contrastive context contained a proposition about an alternative element to evoke a semantic parallelism. The two parts of the semantic parallelism were usually separated by one or two intervening clauses to avoid highly exaggerated realisations. Context-setting in the non-contrastive context was more difficult because we had to ensure that the thematic constituent could be referred to by a full lexical form (instead of pronominalisation). Therefore we decided to resume an initially introduced topic after one or two sentences of elaboration or to use easily inferable relations. The purpose of the production experiment, i.e. the elicitation of identical utterances in pragmatically distinct contexts, was successfully disguised by sufficient unrelated filler paragraphs and by presenting paragraphs on the same topic as the target paragraphs, which did not contain a target sentence.

Our hypothesis was that thematic material would be marked differently, depending on the context it was produced in. Starting from results on contrastive hat patterns in German (Mehlhorn, 2001), we assumed that in contrastive contexts, the initial accent would exhibit a greater f_0-excursion of the rise as well as longer durations for contrastive topics. The most prominent difference between realisations in contrastive and non-contrastive contexts reported by Mehlhorn was the presence of a hat pattern in contrastive contexts. A further introspective difference found in the literature is the depth of the trough before the actual rise, which has been described to be lower in contrastive contexts (Jacobs, 1997). In the acoustic analysis, the annotation of the trough before the actual prenuclear rise turned out to be rather difficult, although we paid careful attention to use sonorant material only. Nevertheless, analysis was hampered, especially by utterance-initial glottalisation.

The results overall showed that themes produced in contrastive contexts had a significantly higher and later peak. In addition, rise excursions were larger in contrastive contexts and rise durations were longer. This coincided with a longer stressed vowel in contrastive contexts. As regards the rheme accent realisation, in contrastive contexts there were significantly more low rheme accents than high ones but in non-contrastive contexts they were equally distributed. Overall, unambiguous hat pattern realisations were not very frequent in production and their use was not a distinctive feature for discriminating contrastive and non-contrastive contexts. The use of significantly more low rheme accents in contrastive contexts suggests, however, that subjects produced some form of “weakened hat pattern”. A particularly interesting finding of the production experiment was that subjects used different “strategies” for signalling contrast: some subjects employed peak height variation almost exclusively (keeping the position of the accent constant), while others relied more on different peak positions (with no or only little variation in peak height). This seemed to indicate a trading relation between peak height and peak alignment. For our (comparatively small) pool of subjects there was no interaction between dialect origin and contrast marking strategy. Further data is needed, however, to isolate possible influences of dialectal origin, which other studies have shown to have an influence on alignment in prenuclear accents (Atterer and Ladd, 2004).

In the perception experiments described in chapter 4 we addressed the question of whether the acoustic differences observed in the production data can be interpreted by listeners. In order to control the phonetic variables involved (especially peak height and peak position, which seemed to be used to a different extent by different speakers), we decided to carry out perception experiments with resynthesised stimuli. Subjects were asked to judge how appropriate a stimulus was in a given contrastive or non-contrastive context. Rating was performed either on a 7-point scale or with magnitude estimation (Bard et al., 1996). Subjects appeared to have less difficulty in using the open scale provided in magnitude estimation compared to the fixed scale.

As contexts for the perception experiments, we deliberately did not use the original contexts of the production study. With these original contexts we would not have been able to isolate the perception of contrast from the perception of discourse structural effects. As contrastive contexts, we set up semantic parallelisms, in which subjects had to judge the appropriateness of the second clause with respect to the visually presented first clause.

Visual presentation:	Anna liebt das Meer, aber ...
	‘*Anna loves the sea but ...*’
Auditory presentation:	Marlene fährt gern in die Berge.
	‘*Marlene likes to go to the mountains.*’

In non-contrastive contexts, subjects were presented a visual pre-context to set the scene. Then they heard a question-answer pair. The question was recorded in a way as to mark the thematic constituent as non-contrastively given ($[_\theta \ldots]$ signals the thematic constituents, $[_\rho \ldots]$ the rhematic ones):

Question:	Was $[_\rho$ MACHT$]$ $[_\theta$ Marlene$]$ denn gerne?
	‘What does $[_\theta$ Marlene$]$ $[_\rho$ like to DO*$]$?’*
Answer:	$[_\theta$ Marlene$]$ $[_\rho$ fährt gern in die BERGE$]$.
	‘$[_\theta$ Marlene$]$ likes $[_\rho$ to go to the MOUNTAINS*$]$.’*

Our hypothesis was that subjects would prefer the stronger marked theme accents (i.e. with a higher or later peak) over less marked ones in contrastive contexts. Guided by our finding of different speaker “strategies” in the production analysis, we assumed that peak height and peak alignment could be additive factors. In non-contrastive contexts we hypothesised the opposite preference pattern, i.e. versions with a later or higher peak should be less preferred. A second resynthesis step manipulated peak height and segmental duration to account for the fact that a later peak could be simply a byproduct of the longer segmental duration. Again we assumed that peak height and segmental duration were additive factors, i.e. in contrastive contexts, subjects were assumed to prefer longer constituents or higher peaks or both. In non-contrastive contexts, we predicted that only short constituents with a low peak would be preferred.

The following discussion first focuses on the results in non-contrastive contexts. These were somewhat unexpected because subjects judged all versions as equally adequate. This result can be explained in two different ways: subjects somehow accommodate a contrast that is not linguistically coded (Thomason, 1990), or they interpret the rising accents not in terms of contrast but make assumptions about the speaker's attitude, his involvement, etc. In order to decide between these alternative explanations, we asked another group of subjects to choose between two stimulus versions in non-contrastive contexts and to explain their choice. It turned out that — in direct comparison — the presumed non-contrastive versions were significantly preferred in the non-contrastive contexts (as previously hypothesised). This choice was usually explained in information structural terms. In those cases in which subjects preferred the presumed *contrastive* versions, they explained their choice with positive emotions or attitudes, such as interest, friendliness, or enthusiasm. This confirms that the prosodic channel can be used to signal different functions. Also, it extends the finding of Nooteboom and Kruyt (1987) who reported that additional accents can be accommodated by listeners; in our case, stronger marked intonational forms (with higher and later peaks) can be accommodated.

In contrastive contexts, listeners had clear preferences. Peak height was a much stronger cue for signalling contrastive themes than peak position was. Versions, which only exhibited a later peak that was not higher pitched were still preferred over versions with an early and low peak but were no substitute for pitch height. This finding undermines the previously assumed trading relation between peak height and peak alignment; it seems that the perception of prenuclear accents is more strongly guided by peak height than by peak alignment. Further, more tailored experiments will be necessary to answer this question decisively. Duration differences, which are usually said to influence prominence relations, did not influence perception at all. Future research will need to tease out the mutual relations between all the variables involved. We are a long way from understanding how these actually interact. Depending on the context, the same intonational marking may have rather different perceptual effects, ranging from the para-linguistic (emotional, attitudinal) to the linguistic (contrast-marking).

In sum, increased pitch height was by far the best cue to signalling thematic contrast. However, pitch height variation is generally interpreted as a gradual difference, which is not perceived categorically (Ladd and Morton, 1997). Therefore we were interested to see whether trained linguists would label the contrastive and non-contrastive version of an utterance pair with a different accent type or with the same one. If they consistently chose a different accent type we could be sure to interpret the differences as categorical. The different accent categories assigned could ascertain phonologically different accent categories, which could be safely used by semanticists to assign thematic contrast. If, however, they mainly used the same accent type to annotate the thematic accent, the differences would have to be interpreted as gradual. Twelve phoneticians, experienced in using GToBI for intonation labelling, were asked to label the prenuclear accents of the two versions of ten of the utterance pairs of our corpus; only seven performed the experiment. The ten utterance pairs for the experiment were randomly selected from the corpus; they had a sufficient difference in peak height and/or peak alignment. The twenty utterances were presented in isolation in random order. For most of the utterance pairs, the prenuclear accents of the two versions were annotated with an identical accent type (note that "identical" does not mean "inter-subjectively identical"; rather, individual labellers annotated the same, probably idiosyncratically preferred prenuclear accent type). There were three utterance pairs, however, for which most of the labellers assigned a different accent type. We could show that the two versions of these utterance pairs mainly differed in the extent of f_0-excursion, which cannot be captured by GToBI categories. That is, if there *are* indeed categorical differences between the two realisations of particular utterance pairs, and if we want to transmit this information about theme accent type to the semantic level, we will have to reconsider the currently available accent types. We will return to this issue below.

6.2 Selected Theoretical Aspects and Future Work

After this brief summary of the experimental results we would like to discuss certain aspects of this thesis that might be taken as thought-provoking impulses or as starting points for further research activity in this area. Some possible extensions to the present experiments have already been proposed in the previous section; a replication of the production study should especially comprise a larger number of subjects of a more balanced dialectal origin. Atterer and Ladd (2004) found that the position of the trough before the prenuclear rise was earlier for Northern German speakers than for Southern German ones. The same trend was reported for the position of the peak. It might be hypothesised that Southern Germans rather use peak height to signal contrast because their peak in neutral realisations is already fairly late. Northern Germans, on the other hand, should be able to use both peak height and peak position to signal contrastive themes. In a similar vein prenuclear material without post-stressed syllables should favour the use of peak height to signal thematic contrast. In previous studies, speech rate variations have helped to pin down segmental anchor points (e.g. Steele, 1986; Ladd et al., 1999); probably this method would shed more light on the issue of whether the prenuclear accents in contrastive and non-contrastive contexts are categorically distinct. Also, it would be of interest to elicit highly exaggerated versions to learn more about the extreme poles. Another focus of future experimentation should involve an investigation of the interaction between the different prosodic variables, such as duration, intensity, and f_0-movement and their respective contribution to the perception of tonal events. In retrospect, our perception experiments with resynthesised stimuli must be seen as rather exploratory, because there turned out to be too many degrees of freedom to thoroughly explore the interaction between all the acoustic variables involved.

Context. In the production experiment we set up two different contexts for a given target utterance: a contrastive one, containing a proposition about an alternative element and a non-contrastive one, in which a high-level topic was resumed. The distinction contrast vs. non-contrast is very focused on specific relations between different utterances, but it ignores other discourse structural relations that might additionally hold between these utterances or discourse relations to other utterances in the larger discourse (Brown and Yule, 1983; Grosz and Sidner, 1986). This underlying hierarchical structure is intrinsically difficult to inspect but it surfaces through the distribution of anaphora and discourse markers on the one hand, and is sometimes signalled prosodically (Swerts, 1994; Wichman et al., 2000). Swerts (1994) discussed the prosodic marking of different degrees of perceived finality, while Wichman et al. (2000) found that the position of the peak differed significantly with the degree of initiality (paragraph initial vs. utterance initial). However, as Swerts pointed out, uncovering and annotating this hierarchical structure is not straightforward: "it is unclear to what extent such an autonomous and reproducible macro-labelling can be achieved on the basis of existing discourse theories" (Swerts, 1997, p. 514). Much more research is necessary to establish operational criteria to reliably annotate the underlying discourse structure. These annotation problems give a rough explanation for the fact that discourse structure formalisation is still in its infancy and why contrast, which seems more traceable, has been treated separately in semantics. On the whole, the distinction contrast vs. non-contrast seems too coarse-grained, however, to capture the possible relations that can hold between utterances. Therefore, contrast should be incorporated on the level of discourse structure (as e.g. done in SDRT, see Asher, 1993), instead of semantics.

The contrastive and non-contrastive contexts we constructed led to prosodically different realisations of the thematic rise. With respect to the perception of these utterances, subjects had clear preferences in contrastive contexts, judging stronger marked rises as more appropriate than less marked ones. In non-contrastive contexts, however, subjects judged all versions as equally appropriate. We could show that listeners seem to make inferences about speaker involvement and interest, i.e. they use information, which is beyond our current definition of context. Kadmon gives a more extensive list of factors that contribute to a given context:

> "Each utterance occurs in a context. This context includes all sorts of things. It includes previous utterances. It includes the speech situation, including the location, the speaker, addressees, various salient objects, and more. It includes various assumptions that the participants in the conversation make about the world in general and about the subject-matter of the conversation in particular. It includes the assumptions that the interlocutors make about the beliefs and intentions of each other. And so on." (Kadmon, 2001, p. 8)

It is clear that these factors cannot be reasonably controlled in experiments. However, much of the above listed "contextual" information is conveyed on the same channel, prosody. This multifunctionality of the prosodic channel makes a simple 1:1 relation from intonational form to meaning very unlikely. It appears, however, that contrastive contexts impose strong constraints on prosodic marking, i.e. subjects judged versions in which certain prosodic variables are missing or too weak as inappropriate. On the other hand, non-contrastive contexts are prosodically underspecified and hence extensive prosodic marking can be interpreted in terms of different functions, such as attitudinal marking. This line of reasoning also accounts for the results reported in Andreeva and Barry (1999), who similarly found that in clearly marked neutral (declarative) contexts, marked forms (questions) were accepted but seen as carrying additional meaning (e.g. impatience or annoyance).

Apart from the linguistic context there is a second sort of contextualisation we would like to address. This second sort is more related to the interpretation of particular accents *in the context* of other intonational phenomena. An example on a global level is the interpretation of the two accents in a hat pattern (cf. section 2.3.3), which — in particularly constructed linguistic examples — is different from the semantics of a lexically identical utterance[1]. An example that is closer to our actual data is the intonational marking of theme and rheme accents. Because of the assumed gradual nature of theme marking, the actual interpretation of thematic material could be facilitated by processing the rheme accent type in combination with it. If the theme accent is sufficiently marked[2] and if the rheme accent is low (H+!H* or H+L*), the probability is fairly high that there is a thematic contrast (even if the contour differs from a prototypical hat pattern). On the other hand, if theme marking is not particularly strong and there is a high rheme accent, the respective theme is probably not contrastive. The same relativity holds for the interpretation of rheme accents: Although there were significantly more low rheme accents than high ones in contrastive contexts, these

[1] In fact, these linguistic examples with hat patterns often involve different accent positions, so that the utterances compared differ not only in accent *type* distribution but also in accent position. Furthermore, the utterances used are more marked linguistic realisations to convey these particular meanings.

[2] According to our data, theme marking has to be a relative notion, which can only be evaluated by comparing the actual accent realisation to a speaker mean.

rheme accent types were equally distributed in non-contrastive context. A low nuclear accent does not necessarily imply a rhematic contrast. If, however, a low rhematic accent coincides with a strongly marked theme accent, a contrast is likely. Again, this model is very crude and we have to investigate many more discourse relations to propose a more refined model. Nevertheless, it seems that the combinatory possibilities of accent types are restricted, possibly in a meaningful way (for a probabilistic grammar of American English accent and phrase accents, as well as boundary tones see Dainora, 2002). Revealing the underlying grammar of accent distributions in utterances seems a promising way to formulate the prosody–semantic interface in a more concise way.

Categoriality and gradience. In the course of our experiments there have been several indications for the view that theme marking is gradiently variable instead of categorically distinct. First, peak height and peak alignment, which are uncorrelated variables, could be used to a different extent to mark contrastive themes. This phenomenon reminded us of trading relations as observed in segmental phonetics. The rating experiments in contrastive contexts revealed that peak height was perceptually more important than mere peak alignment differences, which does not support a trading relation. Nevertheless, these results show that subjects had no difficulty in relating the presumed contrastive realisation to a contrastive context. Since peak height differences, which were the most important cue to the perception of thematic contrast, are usually interpreted as gradually variable (Ladd and Morton, 1997), we additionally conducted a labelling experiment, in which we asked GToBI labellers to annotate the prenuclear accent. We found an annotation task appealing for deciding on the categorical–gradient issue because these labels could have been used as an input to a semantic or pragmatic layer — if the contrastive and non-contrastive versions of the utterance pairs had been consistently annotated with different accent types.

There are several problems with this method, however, which are mainly related to the particular annotation scheme chosen. A major problem of this autosegmental-metrical annotation scheme is that (G)ToBI makes too many theoretical assumptions — which clearly show its phonological roots — to be considered as a useful annotation tool. Apart from such theoretical considerations, (G)ToBI is not able to describe pitch height variations for initial accents, although this pitch height dimension was clearly important for listeners — linguistically naïve ones and annotators. Many of the labellers commented on this very high rise but had no means to annotate it. It should be discussed in the future whether there is a basis for using upstep (ˆH) in intonational phrase intitial rising accents. It is questionable whether other annotation systems would have been theoretically able to capture such peak height differences. American Structuralism, with its four different pitch levels, clearly is, but — for particular reasons — is hardly used anymore. In the annotation schemes developed in the tradition of the British School of intonation, tunes were often distinguished on the basis of the height of the respective start or end points, e.g. *low rise* vs. *high rise*, the former starting from a low pitch, the latter from a medium pitch (O'Connor and Arnold, 1961). The mere presence of categories that account for pitch-excursion difference — which were established on a perceptual basis — indicate the importance of capturing such phenomena as the one under discussion. However, in the particular British School labelling systems, these categories can be only assigned to nuclear tonal movements and not to prenuclear ones. We suggest describing an extra-high pitch excursion in a prenuclear rise with an appropriate diacritic, but we need to clarify first whether this procedure is empirically justified (independent of our data).

With respect to the gradual nature of pitch height distinctions, Ladd argues that it is probably only overall pitch range excursions that have a gradual nature, whereas local pitch range expansions could be seen as categorical:

> "[...] only the *overall* modifications of pitch range are gradient and paralinguistic, and [...] the relative height of accents within phrases and of phrases within sentences is part of the *linguistic* specification of the contour – i.e. part of intonational phonology." (Ladd, 1994, p. 49)

In this article, Ladd proposed to use an "overhigh tone" or "pitch level 4" (the latter notion being reminiscent of the four pitch levels used in American Structuralist approaches) to mark a local pitch range expansion. To back his case for a distinction between high and overhigh, he cites the properties of the language Chaga, an African language, which distinguishes between a "lexical High tone and a contextually raised 'Overhigh' tone" (p. 56). In an earlier article he suggests to use a binary feature "raised peak" (Ladd, 1983), which could be added to particular accents. Experimental results by Nolan (1995) also showed that local emphasis had no influence on the scaling of following accents, which suggests that local "overhigh" pitch excursion can actually be treated as a local phenomenon. Future experimental research has to determine, however, if and how "overhigh" pitch excursion can be operationally defined.

An alternative position is to view prosodic theme marking as gradually variable. From that perspective, a stronger theme marking could be related to increased costs that are necessary to activate themes in different contexts (cf. Chafe, 1994). Activation costs are more fine-grained than the dichotomy contrast vs. non-contrast but are probably of less interest for semanticists (but probably of stronger interest for psychologists).

Assessing On-Line Interpretation

In almost all of our perception experiments, it was implicitly assumed that the ability to correctly relate a resynthesised version to a contrastive or non-contrastive context tells us something about the interpretation of contrastive themes. The only experiment in which we tried to assess interpretation more directly was when we asked subjects to choose between two versions and to explain their choice. Unfortunately, introspective data are not very reliable (see especially the discussion in Nisbet and Wilson, 1977), which calls for methods that allow an on-line inspection of interpretation processes.

Common on-line methods include for instance event-related brain potentials, ERPs[3] (cf. Steinhauer et al., 1999; Hruska et al., 2001), functional magnet-resonance imaging, fMRI[4] (Meyer, 2000; Wildgruber et al., 2001), or eye-tracking (cf. Eberhard et al., 1995; Sedivy, 2003). The neuro-physiological measures are direct only with respect to an immediate reaction of the brain to certain stimuli. The meaning of these neural activity is subject to interpretation of electro-physiological or neuro-spatial data. Therefore, the analysis of eye-movement data (and their relation to auditory stimuli) seems more promising in the near future. Fixations are directly related to what is processed (Eberhard et al., 1995) and can easily be traced. In the recent literature we find a number of eye-tracking experiments that investigated the interpretation of different prosodic structures (Sedivy et al., 1995, 1999; Dahan et al., 2002; Sedivy, 2003), mainly with respect to reference resolution.

Inspired by Sedivy et al. (1999), we conducted an experiment using the visual world paradigm to investigate the influence of contrastive focus, i.e. contrastive rheme accent, on reference resolution (Weber et al., 2004). Instructions to click on a certain object on a computer screen were presented with a high accent on either the adjective or the noun ("Click on the RED comb" vs. "Click on the red COMB"). Four objects were presented on the screen. Two of them were two identical referents with different colours (e.g. a red and a yellow comb, referred to as contrast set); one referent shared the colour with one of these objects, e.g. a yellow ball and one element was an unrelated distractor. In narrow focus (i.e. accent on the adjective), eye-movements to the contrastive object increased immediately after adjective onset, prior to noun onset. Narrow focus hence raised the significance of objects that belonged to a given contrast set (determined by the visual display). In the broad focus condition both the contrastive and non-contrastive object were fixated. This suggests that listeners immediately process contrastive focus to resolve reference ambiguities. These results are promising because they show that interpretation of prosodic information can be monitored using on-line methods. Similarly, Watson et al. (2004b) found that focal H^* and $L+H^*$ accents are processed differently, depending on the visual context. We believe that it is worth exploring this method for prosodic processing further in the future.

---

[3] ERP are electrophysiological data gained by means of electro-encephalograms (EEGs). With this technique, surface electrodes measure brain activity in different cerebral areas during speech perception. Different kinds of neurological activity are associated with different kinds of processing (e.g. lexical, visual or syntactic).

[4] Functional magnet-resonance imaging, which makes use of the fact that activated neural blood vessels carry more oxygen than inactive ones, allows to insect which regions of the brain are active during speech processing. Temporal resolution is still rather poor, however.

Bibliography

Alexopoulou, Theodora. 1996. Contrastive focus and contrastive topics in Greek. *Proc. of the Edinburgh Linguistics Department Conference*.

Allerton, D. J. 1978. The notion of 'givenness' and its relations to presupposition and to theme. *Lingua* **44** pp. 133–168.

Andreeva, Bistra and Tania Avgustinova. 1999. Link-associated accent patterns in Bulgarian. *Proc. of the Third Conference on Formal Description of Slavic Languages (FDSL 3)*, Leipzig.

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

Armstrong, Lilias E. and Ida C. Ward. 1931. *A Handbook of English Intonation*. W. Heffer & Sons, Cambridge.

Arvaniti, Amalia, D. Robert Ladd, and Ineke Mennen. 1998. Stability of tonal alignment: The case of Greek prenuclear accents. *Journal of Phonetics* **26 (1)** pp. 3–25.

Arvaniti, Amalia, D. Robert Ladd, and Ineke Mennen. 2000. What is a starred tone? Evidence from Greek. M. Broe and Janet B. Pierrehumbert (eds.), *Papers in Laboratory Phonology V*, pp. 119–131, Cambridge University Press, Cambridge, UK.

Asher, Nicolas. 1993. *Reference to Abstract Objects in Discourse*. Kluwer Academic Publishers, Dordrecht.

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

Ayers, Gayle M. 1995. Nuclear accent type and prominence: Some psycholinguistic experiments. *Proc. of the 8th International Congress of the Phonetic Sciences*, vol. 3, pp. 660–663, Stockholm, Sweden.

Bach, Kent. 1999. The myth of conventional implicature. *Linguistics and Philosophy* **22 (4)** pp. 327–366.

Bard, Ellen Gurman, Dan Robertson, and Antonella Sorace. 1996. Magnitude estimation of linguistic acceptability. *Language* **72 (1)** pp. 32–68.

Bartels, Christine and John Kingston. 1994. Salient pitch cues in the perception of contrastive focus. Peter Bosch and Rob van der Sandt (eds.), *Focus and Natural Language Processing*, vol. 1 (Intonation and Syntax), pp. 1–10, IBM Deutschland Informationssystem GmbH, Working Paper 6.

Baumann, Stefan and Martine Grice. 2004. Accenting accessible information. *Proc. of the 2nd International Conference on Speech Prosody*, pp. 21–24, Nara, Japan.

Beckman, Mary and Gayle M. Ayers. 1997. Guidelines for ToBI Labelling. *Tech. rep.*, Ohio State University.

Beckman, Mary E. and Janet B. Pierrehumbert. 1986. Intonational structure in Japanese and English. *Phonology Yearbook* **3** pp. 255–309.

Benzmüller, Ralf and Martine Grice. 1997. Trainingsmaterialien zur Etikettierung deutscher Intonation mit GToBI. *Research Report Phonetics, University of the Saarland (PHONUS)* **3** pp. 9–34.

Boersma, Paul. 1992. Praat: A system for doing phonetics by computer. University of Amsterdam.

Bolinger, Dwight L. 1951. Intonation: Levels vs. configurations? *Word* **7** pp. 199–210.

Bolinger, Dwight L. 1961. Contrastive accent and contrastive stress. *Language* **37** pp. 83–96.

Bolinger, Dwight L. 1965. *Forms of English*. Harvard University Press, Cambridge, MA.

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

Braun, Bettina and D. Robert Ladd. 2003. Prosodic correlates of contrastive and non-contrastive themes in German. *Proc. of the 8th European Conference on Speech Communication and Technology*, Geneva, Switzerland.

Braunschweiler, Norbert. 2003. *Automatic Detection of Prosodic Cues*. Ph.D. thesis, University of Konstanz, Germany.

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

Brown, Gilian and George Yule. 1983. *Discourse Analysis*. Cambridge University Press, Cambridge, UK.

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

Bruce, Gösta. 1990. Alignment and composition of tonal accents: Comments on Silverman and Pierrehumbert's paper. John Kingston and Mary E. Beckman (eds.), *Papers in Laboratory Phonology I*, pp. 107–114, Cambridge University Press, Cambridge, UK.

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

Büring, Daniel. 2002. On D-trees, beans, and B-accents, ms.

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

Chafe, Wallace L. 1974. Language and consciousness. *Language* **50** pp. 111–133.

Chafe, Wallace L. 1976. Givenness, contrastiveness, definiteness, subjects, topic, and point of view. C. Li (ed.), *Subject and Topic*, pp. 25–55, Academic Press, New York.

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

Chen, Aoju. 2003. Reaction time as an indicator of discrete intonational contrasts in English. *Proc. of the 8th European Conference on Speech Communication and Technology*, Geneva, Switzerland.

Chomsky, Noam. 1971. Deep structure, surface structure, and semantic representation. D. D. Steinberg and L. A. Jakobovitz (eds.), *Semantics*, pp. 193–217, Cambridge University Press, Cambridge.

Chomsky, Noam and Morris Halle. 1968. *The Sound Pattern of English*. Harper & Row, New York.

Cohen, Antonie and Johan t'Hart. 1967. On the anatomy of intonation. *Lingua* **19** pp. 177–192.

Cohen, Philip, Jerry Morgan, and Martha Pollack (eds.). 1990. *Intentions in Communication*. MIT Press, Cambridge, Massachusetts.

Couper-Kuhlen, Elizabeth. 1986. *An Introduction to English Prosody*. Edward Arnold, London.

Crain, Stephen and Mark Steedman. 1985. On not being led up the garden path: The use of context by the psychological parser. David R. Dowty, Lauri Karttunnen, and Arnold M. Zwicky (eds.), *Natural Language Parsing. Psychological, computational, and theoretical perspectives*, chap. 10, pp. 320–358, Cambridge University Press, Cambridge.

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

Cutler, Anne. 1977. On the context-depence of "intonational meanings". *Papers from the 13th Regional Meeting of the Chicago Linguistic Society*, pp. 104–115.

Cutler, Anne and Jerry A. Fodor. 1979. Semantic focus and sentence comprehension. *Cognition* **7** pp. 49–59.

Dahan, Delphine, Michael K. Tanenhaus, and Craig G. Chambers. 2002. Accent and reference resolution in spoken-language comprehension. *Journal of Memory and Language* **47** pp. 292–314.

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

Dainora, Audra. 2002. Does intonational meaning come from tones or tunes? Evidence against a compositional approach. *Proc. of the 1st International Conference on Speech Prosody*, pp. 235–238, Aix-en-Provence, France.

D'Imperio, Mariapaola. 1999. Tonal structure and pitch targets in Italian focus constituents. *Proc. of the 14th International Congress of the Phonetic Sciences (ICPhS)*, vol. 3, pp. 1757–1760, San Francisco, USA.

Eberhard, Kathleen M., Michael Spivey-Knowlton, Julie C. Sedivy, and Michael K. Tanenhaus. 1995. Eye movements as a window into real-time spoken language comprehension in natural contexts. *Journal of Psycholinguistic Research* **24** pp. 409–436.

Ericcson, K. A. and H. A. Simon. 1980. Verbal reports as data. *Psychological Review* **87** pp. 215–251.

von Essen, Otto. 1964. *Grundzüge der hochdeutschen Satzintonation*. Henn, Ratingen, 2nd edn.

Face, Timothy L. 2002. Local intonational marking of Spanish contrastive focus. *Probus Special Issue on Intonation in Romance Languages* **14** pp. 71–92.

Féry, Caroline. 1993. *German Intonational Patterns*, vol. 285 of *Linguistische Arbeiten*. Niemeyer, Tübingen.

von Fintel, K. 1994. *Restrictions on Quantifier Domains*. Ph.D. thesis, University of Massachusetts, Amherst.

Firbas, J. 1964. On defining the theme in functional sentence perspective. *Travaux Linguistiques de Prague* **1** pp. 267–280.

Fodor, Janet D. 2002. Psycholinguistics cannot escape prosody. *Proc. of the 1st International Conference on Speech Prosody*, pp. 83–88, Aix-en-Provence, France.

Fougeron, Cécile and Pat Keating. 1996. Articulatory strengthening in prosodic domain-initial position. *UCLA Working Papers* **92** pp. 61–87.

Fox, Anthony. 1984. *German Intonation: An Outline*. Clarendon Press, Oxford.

Frota, Sónia. 2002. The prosody of focus: a case-study with cross-linguistic implications. *Proc. of the 1st International Conference on Speech Prosody*, pp. 319–322, Aix-en-Provence, France.

Ginzburg, Jonathan. 1996. Interrogatives: questions, facts, and dialogue. *The Handbook of Contemporary Semantic Theory*, Blackwell Publisher, Oxford, UK.

Goldsmith, John. 1976. *Autosegmetnal Phonology*. Indiana University Linguistics Club.

Grabe, Esther. 2001. IViE labelling guide. web page.
URL `http://www.phon.ox.ac.uk/ esther/ivyweb`

Grabe, Esther. 2002. Variation adds to prosodic typology. *Proc. of the 1st International Conference on Speech Prosody*, pp. 127–132, Aix-en-Provence, France.

Grabe, Esther and Carlos Gussenhoven. 1997. Preaccentual pitch and speaker attitude in Dutch. *Language and Speech* **41 (1)** pp. 63–85.

Grabe, Esther, Francis Nolan, and K. Farrar. 1998. IViE - a comparative transcription system for intonational variation in English. *Proc. of the 5th International Conference on Spoken Language Processing (ICSLP)*, Sydney, Australia.

Grice, Martine and Stefan Baumann. 2002. Deutsche Intonation und GToBI. *Linguistische Berichte* **191** pp. 267–298.

Grice, Martine, Stefan Baumann, and Ralf Benzmüller. 2005. German intonation in autosegmental-metrical phonology. Sun-Ah Jun (ed.), *Prosodic Typology*, pp. 55–83, Oxford University Press, Oxford.

Grice, Martine, D. Robert Ladd, and Amalia Arvaniti. 2000. On the place of phrase accents in intonational phonology. *Phonology* **17 (2)** pp. 143–185.

Grice, Martine, Matthias Reyelt, Ralf Benzmüller, Jörg Mayer, and Anton Batliner. 1996. Consistency in transcription and labelling of German intonation with GToBI. *Proc. of the 4th International Conference on Spoken Language Processin (ICSLP)*, pp. 1716–1719, Philadelphia, USA.

Grimm, C. 1997. Pitch accent in Oneida. *Proc. of the Annual Meeting of the Linguistic Society of America*, Chicago, USA.

Grosz, Barbara and Julia Hirschberg. 1992. Some intonational characteristics of discourse structure. *Proc. of the 2nd International Conference on Spoken Language Processing (ICSLP)*, pp. 429–432, Banff.

Grosz, Barbara J. and Candace L. Sidner. 1986. Attention, intentions, and the structure of discourse. *Computational Linguistics* **12 (3)**.

Gumperz, John J. 1982. *Discourse Strategies*. Studies in Interactional Sociolinguistics, Cambridge University Press, Cambridge.

Gundel, Jeanette and Thorstein Fretheim. 2001. Topic and focus. Laurence Horn and Gregory Ward (eds.), *The Handbook of Pragmatic Theory*, Blackwell, Oxford.

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

Gussenhoven, Carlos. 1983. Focus, mode, and the nucleus. *Journal of Linguistics* **19** pp. 377–417.

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

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

Hadding-Koch, K. and M. Studdert-Kennedy. 1964. An experimental study of some intonation contours. *Phonetica* **11** pp. 175–185.

Hajičová, Eva, Barbara H. Partee, and Petr Sgall. 1998. *Topic-Focus Articulation, Tripartite Structures, and Semantic Context*. Studies in Language and Philosophy, Kluwer Academic Publishers, Dordrecht/Boston/London.

Halliday, Michael A. K. 1967. Notes on transitivity and theme in English. *Journal of Linguistics* **3** pp. 37–81,199–244.

Harnad, Stevan R. (ed.). 1987. *Categorical Perception: The Groundwork of Cognition*. Cambridge University Press, Cambridge.

Hedberg, Nancy. 2002. The prosodic structure of topic and focus in spontaneous English. *Paper at the Workshop on Information Structure in Context*, Stuttgart, Germany.

Heim, Irene R. 1982. *The Semantics of Definite and Indefinite Noun Phrases*. Ph.D. thesis, Amherst, Mass. Univ., published by Garland 88; Outstanding dissertations in linguistics.

Heldner, Mattias. 2003. On the reliability of overall intensity and spectral emphasis as acoustic correlates of focal accents in Swedish. *Journal of Phonetics* **31** pp. 39–62.

Heldner, Mattias and Eva Strangert. 2001. Temporal effects of focus in Swedish. *Journal of Phonetics* **29** pp. 329–361.

von Heusinger, Klaus. 1999. *Intonation and Information Structure*. Habilitationsschrift, Universität Konstanz.

Hirst, Daniel and Albert Di Cristo. 1998. A survey of intonation systems. Daniel Hirst and Albert Di Cristo (eds.), *Intonation Systems: A Survey of Twenty Languages*, pp. 1–44, Cambridge University Press, Cambridge, UK.

Hirst, Daniel, Albert Di Cristo, and Robert Espesser. 2000. Levels of representation and levels of analysis for the description of intonaton systems. *Prosody: Theory and Experiment – Studies Presented to Gosta Bruce*, pp. 51–88, Kluwer Academic Press, Dordrecht, Netherlands.

Hobbs, Jerry. 1990. The Pierrehumbert-Hirschberg theory of intonational meaning made simple: Comments on Pierrehumbert and Hirschberg. Cohen et al. (1990), chap. 15, pp. 313–323.

Hockett, Charles F. 1958. *A Course in Modern Linguistics*. Macmillan, New York.

van Hoof, Hanneke. 2003. The rise in the rise-fall contour: does it evoke a contrastive topic or a contrastive focus. *Linguistics* **41 (3)** pp. 515–563.

House, David. 2003. Perceiving question intonation: the role of pre-focal pause and delayed focal peak. *Proc. of the 15th International Congress of the Phonetic Sciences (ICPhS)*, pp. 755–758, Barcelona, Spain.

Hruska, Claudia, Kai Alter, Karsten Steinhauer, and Anita Steube. 2001. Misleading dialogues: human's brain reaction to prosodic information. Christian Cavé, Isabelle Guaitella, and Serge Santi (eds.), *Oralité et gestualité. Interaction et comportements multimodaux dans la communication. Actes du colloque Orage 2001*, pp. 425–430, L'Harmattan, Paris.

Isačenko, Alexander V. and Hans Joachim Schädlich. 1966. Untersuchungen über die deutsche Satzintonation. *Studia Grammatika* **VII** pp. 7–64.

Jackendoff, Ray. 1972. *Semantic Interpretation in Generative Grammar*. MIT Press, Cambridge, Massachusetts.

Jacobs, Joachim. 1996. Bemerkungen zur I-Topikalisierung. *Sprache und Pragmatik* **41** pp. 1–48.

Jacobs, Joachim. 1997. I-Topikalisierung. *Linguistische Berichte* **108** pp. 91–133.

de Jong, Kenneth. 1994. Initial tones and prominence in Seoul Korean. *Working Papers in Linguistics* **43** pp. 1–14, Ohio State University.

Kadmon, Nirit. 2001. *Formal Pragmatics: Semantics, Pragmatics, Presupposition, and Focus*. Blackwell Publishers, Oxford.

Kamp, Hans. 1981. A Theory of Truth and Semantic Representation. J. A. G. Groenendijk, T. M. V. Janssen, and M. B. J. Stokhof (eds.), *Formal Methods in the Study of Language*, vol. 136, pp. 277–322, Amsterdam: Mathematical Centre. Tracts.

Kamp, Hans and Uwe Reyle. 1993. *From Discourse to Logic: Introduction to Modeltheoretic Semantics of Natural Language, Formal Logic and Discourse Representation Theory*. Studies in Linguistics and Philosophy 42, Kluwer Academic Publishers, Dordrecht/Boston/London.

Karttunen, Lauri. 1974. Presupposition and linguistic context. *Theoretical linguistics* **1** pp. 181–194.

Karttunen, Lauri and Stanley Peters. 1979. Conventional implicature. Choon-Kyu Oh and David A. Dinneen (eds.), *Syntax and Semantics 11. Presupposition*, pp. 1–56, Academic Press, London, New York.

Keller, Frank. 2000. *Gradience in Grammar: Experimental and Computational Aspects of Degrees of Grammaticality*. Ph.D. thesis, University of Edinburgh.
URL `http://www.cogsci.ed.ac.uk/ keller/papers/phd.pdf`

Kim, S.-A. 1999. Positional effect on tonal alternation in Chichewa: phonological rule vs. phonetic timing. *Proc. of the Annual Meeting of of the Chicago Linguistic Society*, vol. 34, pp. 245–257, Chicago, USA.

Kingdon, Roger. 1958. *The Groundwork of English Intonation*. Longman, London.

Kiss, Katalin É. 1998a. Identificational focus vs. information focus. *Language* **74** pp. 245–273.

Kiss, Katalin É. 1998b. Multiple topic, one focus? *Acta Linguistica Hungarica* **45 (1-2)** pp. 3–29.

König, Ekkehard. 1991. *The Meaning of Focus Particles. A Comparative Perspective*. Routledge, London.

Kohler, Klaus. 1997. Modelling prosody in spontaneous speech. Y. Sagisaka, N. Campbell, and N. Higuchi (eds.), *Computing Prosody*, pp. 197–210, 9-Verlag, New York.

Kohler, Klaus J. 1991a. A model of German intonation. *Arbeitsberichte des Instituts für Phonetik der Universität Kiel (AIPUK)* **25** pp. 295–360.

Kohler, Klaus 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** pp. 115–185.

Kohler, Klaus J. and Robert Gartenberg. 1991. The perception of accents: F0 peak height versus F0 peak position. *Arbeitsberichte des Instituts für Phonetik der Universität Kiel (AIPUK)* **25** pp. 219–242.

Kraayefeld, Hans. 1997. *Idiosyncrasy in Prosody. Speaker and Speaker Group Identification in Dutch using Melodic and Temporal Information*. Ph.D. thesis, Universiteit Nijmegen.

Krahmer, Emiel and Marc Swerts. 1999. Contrastive accent in dialogue: Towards a presuppositional analysis. *Proc. ESSLLI Workshop on Focus and Presupposition in Multi-Layer Dialogue*, Utrecht, Netherlands.

Krahmer, Emiel and Marc Swerts. 2001. On the alleged existence of contrastive accents. *Speech Communication* **34** pp. 391–405.

Kratzer, Angelika. 1991. The representation of focus. Armin von Stechow and Dieter Wunderlich (eds.), *Semantik*, vol. 6 of *Handbuch der Sprach- und Kommunikationswissenschaften*, pp. 825–834, Walter de Gruyter, Berlin.

Krifka, Manfred. 1993. Focus and presupposition in dynamic interpretation. *Journal of Semantics* **10 (4)** pp. 269–300.

Kruijff, Geert-Jan M. 2001. *A Categorial-Modal Architecture of Informativity: Dependency Grammar Logic and Information Structure*. Ph.D. thesis, Charles University, Prague, Czech Republic.

Kruijff-Korbayová, Ivana and Mark Steedman. 2003. Two dimensions of information structure in relation to discourse semantics and discourse structure. *Journal of Logic, Language, and Information* **12** pp. 249–259, Special Issue on Information Structure, Disccourse Semantics, and Discourse Structure.

Kuno, Susumo. 1972. Functional sentence perspective: A case study from Japanese and English. *Linguistic Inquiry* **3** pp. 269–320.

Kuno, Susumu. 1993. *The Structure of the Japanese Language*. Current studies in linguistics series, no. 3, MIT Press, Cambrige, MA.

Ladd, D. Robert. 1980. *The Structure of Intonational Meaning*. Indiana University Press, Bloomington.

Ladd, D. Robert. 1983. Phonological features of intonational meaning. *Language* **59** pp. 721–759.

Ladd, D. Robert. 1988. Declination "reset"and the hierarchical organization of utterances. *Journal of the Acoustical Society of America* **84** pp. 530–544.

Ladd, D. Robert. 1994. Constraints on the gradient variability of pitch range, or, Pitch level 4 lives! Patricia A. Keating (ed.), *Papers in Laboratory Phonology III. Phonological Structure and Phonetic form*, pp. 43–63, Cambridge University Press, Cambridge.

Ladd, D. Robert. 1996. *Intonational Phonology*. Cambridge University Press, Cambridge.

Ladd, D. Robert, Dan Faulkner, Hanneke Faulkner, and Astrid Schepman. 1999. Constant "segmental anchoring"of f_0 movements under changes in speech rate. *Journal of the Acoustic Society of America* **106 (3)** pp. 1543–1554.

Ladd, D. Robert, Ineke Mennen, and Astrid Schepman. 2000. Phonological conditioning of peak alignment in rising pitch accents in Dutch. *Journal of the Acoustic Society of America* **107 (5)** pp. 2685–2696.

Ladd, D. Robert and Rachel Morton. 1997. The perception of intonational emphasis: continuous or categorical? *Journal of Phonetics* **25** pp. 313–342.

Ladd, D. Robert and Astrid Schepman. 2003. "Sagging transitions"between high pitch accents in English; experimental evidence. *Journal of Phonetics* **31** pp. 81–112.

Lakoff, George. 1971. Presuppostion and relative well-formedness. Daniel Steinberg and Leon Jakobovits (eds.), *Semantics: An Interdisciplinary Reader in Philosophy, Linguistics and Psychology*, pp. 329–340, Cambridge University Press, Cambridge.

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

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

Lehiste, Ilse. 1970. *Suprasegmentals*. MIT Press, Cambridge MA, London.

Levelt, Willem J.M. 1993. *Speaking: From Intention to Articulation*. MIT Press, Cambridge, Massachusetts.

Levinson, Stephen C. 1983. *Pragmatics*. Cambridge University Press, Cambridge.

Levow, Gina-Anne. 1998. Characterizing and recognizing spoken corrections in human-computer dialogue. *Proc. of COLING-ACL*.

Lewis, David. 1979. Scorekeeping in a language game. *Journal of Philosophical Logic* **8** pp. 339–359.

Liberman, A. M., K. S. Harris, H. S. Hoffman, and B. C. Griffith. 1957. The discrimination of speech sounds within and across phoneme boundaries. *Journal of Experimental Psychology* **61** pp. 379–388.

Liberman, Mark. 1975. *The Intonational System of English*. Ph.D. thesis, MIT.

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

Mann, William C. and S. A. Thompson. 1988. Rhetorical Structure Theory: Toward a functional theory of text organization. *Text* **8 (3)** pp. 243–281.

Mathesius, Vilem. 1929. Functional linguistics. Josef Vachek (ed.), *Praguiana: Some Basic and Less Well-known Aspects of the Prague Linguistics School*, pp. 121–142, John Bejamins, Amsterdam, 1983.

Mehlhorn, Grit. 2001. Produktion und Perzeption von Hutkonturen im Deutschen. *Linguistische Arbeitsberichte* **77** pp. 31–57.

Meyer, Martin. 2000. *Ereignisbezogene hämodynamische Korrelate der Verarbeitung von grammatischer und prosodischer Information*. Ph.D. thesis, University of Leipzig, Germany.

Molnar, Valéria. 2002. Contrast – from a contrastive perspective. H. Hallelgard, Stig Johansson, Bergljot Behrens, and Cathrine Fabricius-Hansen (eds.), *Information Structure in a Cross-Linguistic Perspective*, pp. 147–161, Rodopi, Amsterdam/New York.

Monaghan, A. I. and D. Robert Ladd. 1990. Speaker-dependent and speaker-independent parameters in intonation. *Proc. of the ESCA Workshop on Speaker Characterisation*, pp. 167–174, Edinburgh.

Moulton, William G. 1962. *The Sounds of English and German*. University of Chicago Press, Chicago.

Newmeyer, Frederik J. 2001. The Prague School and North American functionalist approaches to syntax. *Journal of Linguistics* **37 (1)** pp. 101–126.

Nisbet, R. E. and T. D. Wilson. 1977. Telling more than we know: Verbal reports on mental processes. *Psychological Review* **84** pp. 231–259.

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

Nooteboom, Sieb G. and J. G. Kruyt. 1987. Accents, focus distribution, and the perceived distribution of given and new information: An experiment. *Journal of the Acoustical Society of America* **82 (5)** pp. 1512–1524.

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

Oliver, Dominika and Bistra Andreeva. to appear. Peak alignment in broad and narrow focus. *Proc. of 5th European Conference on Formal Description of Slavic Languages*.

Osgood, Charles E., George J. Suci, and Percy H. Tannenbaum. 1957. *The Measurement of Meaning*. University of Illinois Press, Urbana.

Oviatt, Sharon, Gina-Anne Levow, Eliott Moreton, and Margaret MacEachern. 1998a. Modelling global and focal hyperarticulation during human-computer error resolution. *Journal of the Acoustic Society of America* **104 (5)** pp. 3080–3098.

Oviatt, Sharon, Margaret MacEachern, and Gina-Anne Levow. 1998b. Predicting hyperarticulate speech during human-computer error resolution. *Speech Communication* **24 (2)** pp. 1–23.

Palmer, H. E. 1922. *English Intonation, with Systematic Exercises*. Heffer, Cambridge, UK.

Paul, Hermann. 1880. *Principien der Sprachgeschichte/ Sprachgeschichte und Sprachpsychologie*. History of linguistics: 18th and 19th century German linguistics, Routledge/Thoemmes, London, 10th edition, 1995.

Pavlovic, Chaslav V., Mario Rossi, and Robert Espesser. 1990. Use of magnitude estimation technique for assessing the performance of text-to-speech synthesis systems. *Journal of the Acoustic Society of America* **87 (1)** pp. 373–382.

Pavlovic, Chaslav V., Cristel Sorin, J. P. Roumiquière, and J. P. Lucas. 1989. A comparative analysis of the magnitude estimation and the pair comparison techniques for use in assessing qualitiy of text-to-speech synthesis. *Proc. of the ESCA Workshop on Speech I/O Assessment and Speech Databases*, pp. 3.1.1–3.1.3, Noordwijkerhout, Netherlands.

Peters, Jörg, Peter Gilles, Peter Auer, and Margret Selting. 2002. Identification of regional varieties by intonational cues. An experimental study on Hamburg and Berlin German. *Language and Speech* **45 (2)** pp. 115–139.

Pheby, John. 1975. *Intonation und Grammatik im Deutschen*. Akademie-Verlag, Berlin.

Pierrehumbert, Janet. 1980. *The Phonology and Phonetics of English Intonation*. Ph.D. thesis, MIT, Cambridge, Massachusetts, distributed by the IULC.

Pierrehumbert, Janet B. and Julia Hirschberg. 1990. The meaning of intonation in the interpretation of discourse. Cohen et al. (1990), chap. 14, pp. 271–311.

Pierrehumbert, Janet B. and Shirley A. Steele. 1989. Categories of tonal alignment in English. *Phonetica* **46** pp. 181–196.

Pike, Kenneth L. 1945. *The Intonation of American English*. Linguistics 1, University of Michigan Publications, Michigan.

Pisoni, D. B. and J. Tash. 1974. Reaction times to comparisons within and across phonetic categories. *Perception and Psychophysics* **15 (2)** pp. 285–290.

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

Portele, Thomas and Barbara Heuft. 1997. Towards a prominence-based synthesis system. *Speech Communication* **21 (1)** pp. 61–72.

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

Prevost, Scott. 1996. Modelling contrast in the generation and synthesis of spoken language. *Proc. of the 4th International Conference on Spoken Language Processing (ICSLP)*.

Prieto, Pilar, Jan van Santen, and Julia Hirschberg. 1995. Tonal alignment patterns in Spanish. *Journal of Phonetics* **23** pp. 429–451.

Prince, E. 1981. Toward a taxonomy of given/new information. Peter Cole (ed.), *Radical Pragmatics*, pp. 223–254, Academic Press, New York.

Pulman, Stephen G. 1997. Higher order unification and the interpretation of focus. *Linguistics and Philosophy* **20** pp. 73–115.

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

Remijsen, Bert and Vincent J. van Heuven. 1999. Gradient and categorical pitch dimensions in Dutch: Diagnostic test. *Proc. of the 14th International Congress of the Phonetic Sciences (ICPhS)*, pp. 1865–1868, San Francisco, USA.

Repp, B. 1984. Categorical perception: Issues, methods, findings. *Speech and Language: advances in basic research and practice* **10**.

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

Reyelt, Matthias, Martine Grice, Ralf Benzmüller, Jörg Mayer, and Anton Batliner. 1996. Prosodische Etikettierung des Deutschen mit ToBI. Dafydd Gibbon (ed.), *Natural Language and Speech Technology, Results from the third KONVENS conference*, pp. 144–155, Mouton de Gruyter, Berlin, New York.

Rietveld, Toni and Carlos Gussenhoven. 1995. Aligning pitch targets in speech synthesis: effects of syllable structure. *Journal of Phonetics* **23** pp. 375–385.

Roberts, Craige. 1996. Information structure in discourse: Towards an integrated formal theory of pragmatics. *OSU Workshop Papers in Linguistics* **49** pp. 91–136.

Rochemont, Michael S. 1986. *Focus in Generative Grammar*. John Benjamins, Amsterdam/Philadelphia.

Rooth, Mats. 1985. *Association with Focus*. Ph.D. thesis, University of Massachusetts, Amherst.

Rooth, Mats. 1992. A theory of focus interpretation. *Natural Language Semantics* **1** pp. 75–116.

Rossi, Mario. 2000. Intonation: Past, present, future. Antonis Botinis (ed.), *Intonation: Analysis, Modelling and Technology*, Text, Speech, and Language Technology, vol. 15, chap. 8, pp. 13–52, Kluwer, Dordrecht/Boston/London.

Rudolph, Elisabeth. 1996. *Contrast. Adversative and Concessive Expressions on Sentence and Text Level*, vol. 23 of *Research in Text Theory. Untersuchungen zur Texttheorie*. de Gruyter.

Sag, Ivan and Mark Liberman. 1975. The intonational disambiguation of indirect speech acts. *Papers from the 11th Regional Meeting of the Chicago Linguistic Society*, pp. 487–497.

Schmerling, Susan Fred. 1976. *Aspects of English Sentence Stress*. University of Texas Press, Austin, Texas.

Schouten, Bert, Ellen Gerrits, and Arjan van Hessen. 2003. The end of categorical perception as we know it. *Speech Communication* **41** pp. 71–80.

Sedivy, Julie C. 2003. Pragmatic versus form-base accounts of referential contrast: Evidence for effects of informativity expectations. *Journal of Psychological Research* **32 (1)** pp. 3–24.

Sedivy, Julie C., Michael K. Tanenhaus, Kathleen Eberhard, Michael Spivey-Knowlton, and Gregory N. Carlson. 1995. Using intonationally-marked presuppositional information in on-line language processing: Evidence from eye movements to a visual model. *Proc. of the 17th Annual Conference of the Cognitive Science Society*, pp. 375–380, Hillsdale, NJ.

Sedivy, Julie C., Michel K. Tanenhaus, Craig G. Chambers, and Gregory N. Carlson. 1999. Achieving incremental semantic interpretation through contextual interpretation. *Cognition* **71** pp. 109–147.

Selkirk, Elisabeth. 1984. *Phonology and Syntax: The Relation between Sound and Structure*. MIT Press, Cambridge, Massachusetts.

Selkirk, Elisabeth. 1995. Sentence prosody: Intonation, stress, phrasing. John A. Goldsmith (ed.), *Handbook of Phonological Theory*, Blackwell, Cambridge.

Selkirk, Elisabeth. 2002. Contrastive FOCUS vs. presentational *focus*: Prosodic evidence from right node raising in English. *Proc. of the 1st International Conference on Speech Prosody*, pp. 643–646, Aix-en-Provence, France.

Shriberg, Elizabeth, Rebecca Bates, Andreas Stolcke, Paul Taylor, Daniel Jurafsky, K. Ries, N. Coccaro, R. Martin, M. Meteer, and C. Van Ess-Dykema. 1998. Can prosody aid the automatic classification of dialog acts in conversational speech? *Language and Speech* **41 (3–4)** pp. 439–487.

Shriberg, Elizabeth, Andreas Stolcke, Dilek Hakkani-Tür, and Gökhan Tür. 2000. Prosody-based automatic segmentation of speech into sentences and topics. *Speech Communication* **32** pp. 127–154.

Silverman, Kim E. A., Mary E. Beckman, John Pitrelli, Mari Ostendorf, Colin Wightman, Patti Price, Janet B. Pierrehumbert, and Julia Hirschberg. 1992. ToBI: A standard for labeling English prosody. *Proceedings of the 2nd International Conference on Spoken Language Processing*, vol. 2, pp. 867–870, Banff, Canada.

Silverman, Kim E. A. and Janet B. Pierrehumbert. 1990. The timing of prenuclear high accents in English. John Kingston and Mary E. Beckman (eds.), *Papers in Laboratory Phonology I*, pp. 72–106, Cambridge University Press, Cambridge, UK.

Sluijter, Agaath M. C. and Jacques M. B. Terken. 1993. Beyond sentence prosody: Paragraph intonaton in Dutch. *Phonetica* **50** pp. 180–188.

Sluijter, Agatha M.C. 1995. *Phonetic Correlates of Stress and Accent*. ICG Printing, Dordrecht, Proefschrift Rijksuniversiteit Leiden.

Sperber, Dan and Deirdre Wilson. 1982. Mutual knowledge and relevance in theories of comprehension. Neil Smith (ed.), *Mutual Knowledge*, pp. 61–68, Academic Press, New York.

von Stechow, Armin. 1990. Focusing and backgrounding operators. W. Abraham (ed.), *Discourse Particles*, John Benjamins, Amsterdam.

von Stechow, Arnim. 1991. Current issues in the theory of focus. Armin von Stechow and Dieter Wunderlich (eds.), *Semantik*, vol. 6 of *Handbuch der Sprach- und Kommunikationswissenschaften*, pp. 804–825, Walter de Gruyter, Berlin.

Steedman, Mark. 1991. Structure and intonation. *Language* **67 (2)** pp. 260–296.

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

Steedman, Mark. 2002. Information-structural semantics of English intonation, paper to LSA Summer Institute Workshop on Topic and Focus, Santa Barbara, 2001.

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**.

Steinhauer, Karsten, Kai Alter, and Angela D. Friederici. 1999. Brain potentials indicate immediate use of prosodic cues in natuaral speech processing. *Nature Neuroscience* **2** pp. 191–196.

Steube, Anita. 2001. Grammatik und Pragmatik von Hutkonturen. *Linguistische Arbeitsberichte* **77** pp. 7–29.

Stevens, S. S. 1957. On the psychophysical law. *Psychological Review* **64** pp. 153–181.

Strawson, Peter F. 1950. On referring. *Mind* **59** pp. 320–344.

Strawson, Peter F. 1964. Identifying reference and truth-values. Daniel Steinberg and Leon Jacobovits (eds.), *Semantics*, pp. 86–99, Cambridge University Press.

Streefkerk, Barbertje M., Louis C. W. Pols, and Louis F. M Ten Bosch. 1998. Automatic detection of prominence (as defined by listners' judgements) in read aloud Dutch sentences. *Proc. of the 5th International Conference on Spoken Language Processing (ICSLP).*

Street, Richard L. 1982. Evaluation of noncontent speech accommodation. *Language and Communication* **2 (1)** pp. 13–31.

Sweet, Henry. 1878. *Handbook of Phonetics*. Clarendon Press, Oxford.

Sweet, Henry. 1892. *A Primer in Phonetics*. Clarendon Press, Oxford.

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

Swerts, Marc. 1997. Prosodic features at discourse boundaries of different strength. *Journal of the Acoustic Society of America* **101 (1)** pp. 514–521.

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

Tanenhaus, Michael K., Michael Spivey-Knowlton, Kathleen M. Eberhard, and Julie C. Sedivy. 1995. Integration of visual and linguistic information in spoken language comprehension. *Science* **268** pp. 1632–1634.

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

Terken, Jacques and Julia Hirschberg. 1994. Deaccentuation of words representing 'given' information: Effects of persistence of grammatical function and surface position. *Language and Speech* **37 (2)** pp. 125–145.

Terken, Jacques M. B. 1985. *Use and Function of Accentuation: Some Experiments*. Ph.D. thesis, University of Leiden.

t'Hart, Johan and Antonie Cohen. 1973. Intonation by rule: a perceptual quest. *Journal of Phonetics* **1** pp. 309–327.

t'Hart, Johan, René Collier, and Antonie Cohen. 1990. *A Perceptual Study of Intonation: An Experimental-Phonetic Approach*. Cambridge University Press, Cambridge.

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

Thomason, Richmond H. 1990. Accomodation, meaning, and implicature: Interdisciplinary fondations for pragmatics. Cohen et al. (1990), chap. 16, pp. 325–363.

Timm, Uwe. 2003. *Rot*. Deutscher Taschenbuch Verlag.

Toepel, Ulrike and Kai Alter. 2003. How mis-specified focus accents can distress our brain. *Proc. of the 15th International Congress of the Phonetic Sciences (ICPhS)*, pp. 619–622, Barcelona, Spain.

Trager, George L. 1964. The intonation system of American English. David Abercrombie et al. (ed.), *In Honour of Daniel Jones*, pp. 266–270, Longman, London.

Trager, George L. and Henry L. Smith. 1957. *An Outline of English Structure*. Studies in Linguistics: Occasional Papers 3, American Council of Learned Societies, Washington D.C., 5th edn.

Trubetzkoy, Nikolay. 1939. *Grundzüge der Phonologie*. Vandenhoeck and Ruprecht, Göttingen.

Turk, Alice and Laurence White. 1999. Structural influences on accentual lengthening in English. *Journal of Phonetics* **27** pp. 171–206.

Umbach, Carla. 2004. On the notion of contrast in information structure and discourse structure. *Journal of Semantics* **21 (2)**.

Vallduví, Enric. 1990. *The Informational Component*. Ph.D. thesis, University of Pennsylvania, Philadelphia, USA.

Vallduví, Enric and Elisabeth Engdahl. 1996. The linguistic realisation of information packaging. *Linguistics* **34** pp. 459–519.

Vallduví, Enric and Maria Vilkuna. 1998. On rheme and kontrast. Peter Culicover and Louise McNally (eds.), *The Limits of Syntax*, vol. 29 of *Syntax and Semantics*, pp. 79–108, Academic Press, New York, San Francisco, London.

Vallduví, Enric and Ron Zacharski. 1994. Accenting phenomena, association with focus, and the recursiveness of focus–ground. P. Dekker and M. Stockhof (eds.), *Proc. of the 9th Amsterdam Colloquium*, pp. 683–702, ILLC, Amsterdam.

Välimaa-Blum, Ritta. 1988. *Finnish Existential Clauses: their Syntax, Pragmatics, and Intonation*. Ph.D. thesis, Ohio State University.

von Stechow, Arnim and Susanne Uhmann. 1986. Some remarks on focus projection. Werner Abraham and S. de Meij (eds.), *Topic, Focus, and Configurationality. Papers from the 6th Groningen Grammar Talks, Groningen, 1984*, no. 4 in Linguistik Aktuell / Linguistics Today, pp. 295–320, John Benjamins Publishing Company, Amsterdam, Philadelphia.

Ward, Gregory and Julia Hirschberg. 1985. Implicating uncertainty: The pragmatics of fall-rise intonation. *Language* **61** pp. 747–776.

Watson, Duane G., Michael K. Tanenhaus, and Christine A. Gunlogson. 2004a. Interpreting ptich accents in on-line comprehension: H* vs L+H*, under review.

Watson, Duane G., Michael K. Tannenhaus, and Christine A. Gunlogson. 2004b. Processing pitch accents: Interpreting H* and L+H*. *Poster presented at the 17th Annual CUNY Sentence Processing Conference*, CUNY, Boston, USA.

Weber, Andrea, Bettina Braun, and Matthew W. Crocker. 2004. Adjectival modifiers and reference resolution: When prosodic focus matters. *Poster presented at the 17th Annual CUNY Sentence Processing Conference*, CUNY, Boston, USA.

Wells, William H. B. 1988. *Focus in Spoken English*. Ph.D. thesis, University of York.

Wichman, Anne. 2000. *Intonation in Text and Discourse*. Pearson Education, London, UK.

Wichman, Anne, Jill House, and Toni Rietveld. 2000. Discourse constraints on F_0 peak timing in English. Antonis Botinis (ed.), *Intonation: Analysis, Modelling and Technology*, Text, Speech, and Language Technology, vol. 15, chap. 8, pp. 163–182, Kluwer, Dordrecht/Boston/London.

Wightman, Colin W. and Mari Ostendorf. 1994. Automatic labelling of prosodic patterns. *IEEE Transactions on Speech and Audio Processing* **2 (4)** pp. 469–481.

Wildgruber, Dirk, Ingo Hertrich, Hermann Ackermann, Axel Riecker, and Wolfgang Grodd. 2001. Processing of linguistic and affective aspects of speech intonation evaluated by fmri. *NeuroImage* **13 (6)**.

Williams, Carl E. and Kenneth N. Stevens. 1972. Emotions and speech: Some acoustic correlates. *Journal of the Acoustic Society of America* **52** pp. 1238–1250.

Wunderlich, Dieter. 1988. Der Ton macht die Melodie – Zur Phonologie der Intonation des Deutschen. Hans Altmann (ed.), *Intonationsforschungen*, Linguistische Arbeiten 200, pp. 1–40, Niemeyer, Tübingen.

Wunderlich, Dieter. 1991. Intonation and contrast. *Journal of Semantics* **8** pp. 239–251.

Xu, Yi. 1998. Consistency of tone-syllable alignment across different syllable structures and speaking rates. *Phonetica* **55** pp. 179–203.

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

Zubizarreta, Maria Luisa. 1998. *Prosody, Focus, and Word Order*. Linguistic Inquiry, MIT Press, Cambridge, Massachusetts.

Appendix A

Production Experiment

A.1 Dialectal Origin of Speakers

speaker	origin	northern	sex
BK	Lower Saxony	1	female
CS	Lower Saxony	1	female
UB	Lower Saxony	1	female
SZ	North Rhine-Westfalia	1	female
IB	Baden-Wuerttemberg	0	female
CH	Bavaria	0	female
PK	Bavaria	0	female
JL	Palatinate	0	male
CZ	Hesse	1	male
MB	Berlin	1	male
DS	North Rhine-Westfalia	1	male
SV	Schleswig-Holstein	1	male

Table A.1: Information about the speakers

A.2 Stimuli

The stimulus pairs are grouped according to the grammatical relation of the constituent in the preverbal field. The target utterance is indicated in bold face. The first paragraph of a pair establishes the non-contrastive context, the second paragraph the contrastive one.

Fronted subj-NP (agentive)

1. In diesem Jahr haben wir eine neue Mitarbeiterin bekommen. Sie heißt Marlene und kommt ursprünglich aus Berlin. Wir verstehen uns schon alle recht gut mit ihr. Sie

ist eine freundliche und ausgeglichene Person und hat einige besondere Fähigkeiten. **Marlene spielt Klavier und kann singen**. Auf unserer nächsten Institutsfeier ist sie damit sicher der Star.

In diesem Jahr haben wir zwei neue Kolleginnen bekommen, Marlene und Anna. Sie sind beide sehr nett und wir verstehen uns ausgezeichnet mit ihnen. Die beiden sind schon länger sehr gut befreundet, auch wenn sie sehr unterschiedlich sind. **Marlene spielt Klavier und kann singen**. In ihrer Freizeit schreibt sie außerdem Gedichte. Anna ist mehr der sportliche Typ, der auch mal eine Nacht in einer Disko zubringt.

2. Unser Nachbar hat mindestens 20 Lämmer. Sie sehen total süß aus mit ihrem wuscheligem Fell und den treuherzig wirkenden Augen. Mich überkommt jedesmal das Bedürfnis, sie zu streicheln. Aber der Bauer hat mir erklärt, dass das nicht geht. **Die Lämmer haben Angst vor Menschen**. Und nicht einmal ausgewachsene Schafe trauen sich zu nah an die Menschen heran.

   Unser Nachbar hat sehr viele Tiere, vor allem Jungtiere. Ich war als Kind oft bei ihm und habe ihm geholfen, die Tiere zu versorgen. Fohlen haben sich gut an Menschen gewöhnt. Sie fressen ihnen im wahrsten Sinne des Wortes ‘aus der Hand’. **Die Lämmer haben Angst vor Menschen**. Ihnen kann man sich nicht auf 2 Meter nähern, sonst laufen sie weg.

3. 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.

   ⇒ “Malayen” actually refers to the ethnic group whereas the correct German term for the people who live in Malaysia would be “Malaysier”. Because of its excellent sonorant structure, we used the *wrong* term.

4. Lilienzwiebeln werden nach ihrer Blütezeit gepflanzt, d.h. im August oder September. Der heimische Standort vieler Lilienarten sind Bergniederungen oder niedrige Gebirgslagen. **Lilien bevorzugen steinigen Boden**. Dieser kann mit Humus bedeckt werden. Man sollte dafür sorgen, dass die Lilien ausreichend mit Wasser versorgt werden.

Lilien und Rosen spielten in der Dichtung schon immer eine große Rolle. Noch heute gelten diese beiden Arten als besonders edel. Bestimmte Lilienarten blühen bereits im März, während die Rose meist erst im Juni in voller Blüte steht. **Lilien bevorzugen steinigen Boden**. Sie haben kein Problem mit hohen Niederschlagsmengen. Rosen sind weitaus empfindlicher und müssen regelmäßig gedüngt werden.

5. Die Libellen gelten als die größten Insekten. Sie können eine Flügelspannweite von bis zu 20 cm entwickeln. Bei einheimischen Tieren liegt das Maximum bei 10 cm. **Libellen bevorzugen Sumpfgebiete**. Obwohl sie wegen ihrer großen Augen auf die Menschen bedrohlich wirken, sind sie vollkommen harmlos.

   Viele Menschen haben Angst vor Libellen. Insbesondere haben sie Angst, gestochen zu werden. Diese Verwechslung mit Mücken kommt leider oft vor, obwohl diese beiden Arten nichts gemein haben. Mücken halten sich in der Nähe von Menschen auf und leben von Blut. **Libellen bevorzugen Sumpfgebiete**. Außerdem halten sie sich von Menschen fern.

6. Die Alemannen waren ein germanisches Volk, das den Römern am Oberrhein gegenüberstand. Sie wurden um 500 nach Christus dem Frankenreich eingegliedert, was ihnen viele Rechte nahm. **Die Alemannen verfügten über kein Thronfolgerecht**. Die jeweiligen neuen Herzöge wurden von den Merowingern eingesetzt. Erst 700 gelang es den Alemannen, ihr Herzogtum fest zu etablieren.

   Im Frankenreich waren die Merowinger ein mächtiger Faktor. Ihnen gelang es, über Jahrhunderte hinweg ihre Macht zu erhalten. Andere Völker wurden geschickt unterdrückt. Die Bajuwaren z.B. durften ihre Schätze nicht ungeteilt vererben. **Die Alemannen verfügten über kein Thronfolgerecht**. Ihre Herzöge wurden von den Merowingern ein- und abgesetzt.

7. Während meines Studiums habe ich ein Praktikum in Norwegen gemacht. Ich war für 8 Monate bei einer Computerfirma in Oslo und habe dort programmiert. Mit den Arbeitskollegen hab ich mich prima verstanden und ich hab sie auch in der Freizeit oft getroffen. **Die Norweger geben gern Parties**. So hab ich ziemlich viele Leute kennengelernt.

   Ich bin seit drei Jahren mit einem Norweger verheiratet. Auch wenn wir uns schon sehr gut kennen, bemerke ich immer noch kulturelle Unterschiede zwischen uns. Ich, als Deutsche, bin gern abends mit meinem Mann zuhause. Ich genieße es, zu zweit auf dem Sofa zu liegen. **Die Norweger geben gern Parties**. Irgendwie scheinen sie den Rummel zu brauchen.

8. Die offizielle Bezeichnung für die “Mormonen” ist “Kirche Jesu Christi der Heiligen der Letzten Tage”. Diese Religionsgemeinschaft wurde im 19. Jahrundert in Ameri-

ka gegründet. Der Hauptsitz ist immer noch in Utah. **Die Mormonen sind sehr traditionell**. Das wichtigste Element ist die Familie.

Während bei den christlichen Kirchen die Mitgliederzahl stark zurückgeht, nimmt die Gemeinschaft der Mormonen ständig zu. Ein häufig genannter Grund ist, dass katholische und protestantische Gemeinden zu locker und zu freizügig sind. **Die Mormonen sind sehr traditionell**. Ihre strengen Glaubensgrundsätze sind für viele Menschen anziehend.

Existential Sentence

1. Dieses Jahr war ich mit Freunden auf den Kanaren im Urlaub. Wir haben nur den Flug gebucht und haben uns vor Ort auf eigene Faust durchgeschlagen. Teneriffa und La Gomera haben uns besonders gut gefallen. Wir sind viel am Strand gelegen, haben aber auch längere Wanderungen im Landesinneren unternommen. **Die Kanaren sind ein Wanderparadies**. Nächstes Jahr fahre ich bestimmt wieder hin.

   Meine letzten Urlaubsreisen gingen immer zu Inseln. Letzes Jahr war ich auf den Balearen und dieses Jahr auf den Kanaren. Die Balearen laden besonders zu Wassersport ein. Ich haben dort Surfen gelernt und bin geschnorchelt. **Die Kanaren sind ein Wanderparadies**. Ich habe oft Touren unternommen, die 5 bis 6 Stunden gedauert haben.

2. Wir verbringen unseren Urlaub oft in Italien. Da wir nicht gern lange im Auto sitzen, fahren wir meist nach Norditalien. Unsere Lieblingsregionen sind die Toskana und Umbrien. Neben der Gegend fahren wir vor allem auch wegen der Menschen hin. **Italiener sind sehr gastfreundlich**. Wir haben dort schon viele Freunde gewonnen, die uns auch manchmal in Deutschland besuchen kommen.

   Andere Länder, andere Sitten. Oder ist vielleicht auch die Mentalität, die Empfindung eine andere? Dazu ein einfaches Beispiel: Deutsche leben gern für sich. Sie sind froh, in ihren vier Wänden unter sich zu sein. **Italiener sind sehr gastfreundlich**. Man hat das Gefühl, sie wollen die Gäste gar nicht mehr gehen lassen. Vielleicht ist es auch nur Konvention.

3. Ich habe mich schon immer für Geschichte interessiert, vor allem für die Römer. Mich fasziniert besonders, wie sie es schafften, so schnell ihr Reich zu vergrößern und fast ganz Europa zu regieren. **Die Römer waren sehr organisiert**. Dies hat ihnen bei ihrem Feldzug durch Europa sicher geholfen.

   Die Germanen haben verzweifelt versucht, sich gegen die Römer zur Wehr zu setzen, sind aber letztlich doch unterlegen. Das hatte nicht nur mit der Truppenstärke zu tun. Die Germanen waren in sich gespalten und es gab keinen allgemein anerkannten

Anführer. **Die Römer waren sehr organisiert**. Außerdem waren sie es gewohnt, Befehle jeder Art auszuführen.

4. Vermutlich wissen nicht mehr viele Kinder, wie es ist, selbst Marmelade zu machen. Je nach Fruchtsorte dauert schon das Sammeln der Früchte ziemlich lange. Wenn aber erst der süße Geruch durch die Wohnung zieht, ist das Lohn genug für die Anstrengung. Ernährungstechnisch gilt sie leider nicht als vollkommen gesund. **Marmelade ist sehr zuckerhaltig**. Diese teilweise negative Bilanz wird aber vermutlich an deren Beliebtheit nichts ändern.

   Die Sorge um die richtige Ernährung des Nachwuchses wird für Eltern zunehmend wichtiger. Da Kinder aber durchaus ihren eigenen Geschmack haben, ist es nicht immer einfach, nur gesundes Essen in die Kleinen zu füttern. Der Stress beginnt meistens schon mit dem Frühstück und dem 'richtigen' Brotaufstrich. Schinken und Käse sind bei den Kindern oft nicht sehr beliebt. Nutella ist beliebt, aber sehr fett. **Marmelade ist sehr zuckerhaltig**. Das Müsli wäre wohl die beste Alternative.

5. Der Beruf des Malers ist sehr vielseitig. Neben Grundkenntnissen über die verschiedenen Materialien ist auch handwerkliches Geschick gefragt. Sie schützen und schmücken Decken, Wände und Hausfassaden. Besonders interessant ist die Sanierung von Altbauten. **Die Maler sind viel im Freien.** Daher brauchen sie eine gute körperliche Verfassung.

   Die Berufe "Maler" und "Gipser" sind bei vielen Jugendlichen nicht sehr beliebt. Unter den Handwerksberufen sind Schreiner, Schlosser und Maurer die beliebtesten. Gipser arbeiten meist in Wohnungen und beschäftigen sich unter anderem mit der Reparatur von Stuckwerk. **Die Maler sind viel im Freien**. Für den Maler ist es besonders wichtig, dass er ein Gespür für Farben hat.

Phrase-Initial Locative and Temporal Prepositional Phrases

1. Letzte Woche hat meine Freundin mit der Urlaubsplanung für dieses Jahr begonnen. Sie ist Lehrerin in Bayern und daher auf die staatliche Ferienplanung angewiesen. Ihre momentanen Pläne sind, eine Woche in der Sahara zu verbringen, um mal so etwas ganz anderes zu erleben. **In Bayern beginnen die Ferien Ende Juli**. Das könnte dann für die Sahara fast ein bisschen zu heiß werden.

   In Deutschland sind die Ferienzeiten nicht so einheitlich geregelt wie in Frankreich. Aus diesem Grund können meine Freunde und ich, die wir alle Kinder haben und in verschiednen Bundesländern leben, nie gemeinsam in Urlaub fahren. **In Bayern beginnen die Ferien Ende Juli**. Das ist die Zeit, wo für Saarländer das neue Schuljahr beginnt.

2. Letztes Jahr hab ich meinen Urlaub in Mailand verbracht, oder ‘Milano’, wie die Italiener dazu sagen. Obwohl die Stadt hauptsächlich von Industrie lebt, hat sie doch einen gewissen Reiz. An Samstagen wird es in der Innenstadt besonders voll. Und es gibt einen guten Grund, warum so viele Leute anreisen. **In Milano kann man gut einkaufen**. Das bezieht sich vor allem auf italienische Schuhe und die neueste Damenmode.

   Ein wassersportbegeisterter Freund hat mich gefragt, ob es in der Nähe des Gardasees auch für seine Freundin was Interessantes gäbe. Kulturell gesehen ist Verona die sehenswerteste Stadt. Aber es gibt noch weitere interessante Städte, die zu einem Besuch einladen. Wer guten Schinken kaufen möchte, der ist in Parma gut aufgehoben. **In Milano kann man gut einkaufen**. Nördlich des Gardasees ist Trento mit seinem mittelalterlichen Stadtbild ungeschlagen.

3. Letztes Jahr hatte ich im Januar eine australische Freundin zu Besuch. Es hatte bereits kräftig geschneit und die ganze Landschaft war weiß. **Im Januar ist es frostig**. Für meine Freundin von der Südhalbkugel war das besonders schlimm, weil sie ja praktisch aus dem Sommer kam. Der Schnee hat ihr aber trotzdem gut gefallen und wir waren oft draußen.

   Eine australische Freundin, die über Weihnachten nach Europa kommt, hat sich nach der Temperatur in den Wintermonaten erkundigt. Der Dezember ist in Deutschland ja vergleichsweise mild. Es kommt immer seltener vor, dass es an Weihnachten schneit. **Im Januar ist es frostig**. Dieses Jahr waren z.B. -20 Grad Celsius in der Nacht keine Seltenheit. Im Februar kann es schon einige frühlingshafte Tage geben.

4. Amerika gilt immer noch als das ‘Land der unbegrenzten Möglichkeiten’. Der Staat hält sich weitgehend aus den Angelegenheiten seiner Bürger raus, d.h. Selbstverantwortung steht im Vordergrund. Dies bezieht sich nicht nur auf die Sozialgesetzgebung und die Bildung. **In Amerika besitzen viele eine Waffe**. Dies gehört ebenso zum persönlichen Schutz wie eine Alarmanlage.

   In Deutschland hat der Amoklauf eines Schülers wieder einmal die Diskussion entfacht, ob Schusswaffen generell verboten werden sollten. Während hierzulande eine breite Bevölkerungsschicht dem Besitz von Waffen kritisch gegenübersteht, gibt es solche Vorbehalte in anderen Ländern nicht. **In Amerika besitzen viele eine Waffe**. Der Waffenbesitz ist in vielen Staaten so natürlich wie der Personalausweis.

5. Armenien ist eine ehemalige sowjetische Republik, die 1988 unabhängig wurde. Während die Armenier innerhalb der Sowjetunionn zu relativem Wohlstand kamen, leben heute viele in Armut. Armenien ist ein Binnenland und wird von ca. 2 Milionen Menschen bewohnt. Ungefähr 80% der Bevölkerung sind Christen. **In Armenien**

schreibt man lateinisch. Diese westliche Orientierung ist sehr ungewöhnlich für den Kaukasus.

In den ehemals sowjetischen Ländern Georgien, Armenien und Aserbaidschan gibt es viele Konflikte. Im Vordergrund stehen territoriale Interessen, aber diese Länder sind insgesamt grundverschieden. In Aserbaidschan sind viele pro-russisch. Die Georgier hingegen besitzen sogar eine eigene Schrift. **In Armenien schreibt man lateinisch**. Es müssen also erst sprachliche Hindernisse überwunden werden, um in einen Dialog zu treten.

⇒ it is not true that Armenia uses the Latin alphabet. This was pretended for the sake of a semantically coherent paragraph.

A.3 Fillers

Genuine Filler Paragraphs

1. Am Freitagnachmittag war ich bei einem Freund zum Kaffee eingeladen. Er ist erst kürzlich in eine neue Wohnung in Frankfurt gezogen. Er hat eine 2-Zimmerwohnung mit einem riesigen Wohnzimmer. Die Wohnung ist sehr hell, weil sie im achten Stock liegt. Die umliegenden Gebäude sind alle ein bisschen niedriger.

2. Annabell ist eine neue Studentin bei uns am Institut. Sie ist Amerikanerin, aber ihre Eltern kommen eigentlich aus Frankreich. Annabell ist in Frankreich geboren und ich denke, dass sie auch ihre ersten Schuljahre dort verbracht hat. Ihr Englisch ist trotzdem ausgezeichnet, und man würde wohl nicht vermuten, dass sie eigentlich Französin ist.

3. Die Mietpreise sind in den letzten Jahren drastisch angestiegen. Diese Entwicklung ist nicht nur auf die beliebten Großstädte wie z.B. München und Berlin beschränkt. Betroffen sind mehr und mehr auch kleinere Städte. Ein Grund für diese Entwicklung ist sicherlich der steigende Anteil an Single-Haushalten.

4. Meine Freundin studiert momentan in England, genauer gesagt in London. Sie hat sich schon im zweiten Semester intensiv um Stipendien bemüht und ist jetzt nach dem Vordiplom umgezogen. Eigentlich wollte sie nie in einer Metropole leben. Ich habe ihr aber geraten, es einfach mal auszuprobieren, weil man so eine Gelegenheit nicht oft hat. Außerdem ist sie noch jung und aufgeschlossen und daher sicher in der Lage, sich in die neue Situation einzufinden.

5. Es scheint nach wie vor ungeklärt, ob der *mäßige* Konsum von Alkohol und vor allem von Rotwein positive oder negative Auswirkungen auf den menschlichen Organismus hat. Wissenschaftliche Studien kommen zu sehr unterschiedlichen Ergebnissen. Während auf der einen Seite vor dem erhöhtem Krebsrisiko durch Alkoholgenuss gewarnt wird, zeigen andere Studien positive Effekte auf das Herz-Kreislauf-System.

6. Es ist interessant zu beobachten, wie sich die Wellness-Bewegung und die Fast-food-Bewegung parallel entwickeln. Während sich auf der einen Seite immer mehr Menschen für gesunde Ernährung, ausreichend Bewegung und Wellness-Wochenenden entscheiden, wächst gleichzeitig der Anteil an Menschen, die sich überhaupt nicht für Gesundheit interessieren. Einen Mittelweg scheint es kaum mehr zu geben.

7. Als Kinder haben wir viel im Freien gespielt. Manchmal haben wir uns geweigert, weil wir gerne weiter im Zimmer mit Lego spielen wollten. Aber sobald es draußen trocken war, wurden wir von unseren Eltern rausgeschickt. Vermutlich haben wir gemacht, was alle Kinder unternehmen: Fußball spielen, schaukeln, Sandburgen bauen oder Tischtennis spielen. Im nachhinein betrachtet war es eine wunderschöne Zeit.

8. Der neue Roman von Martin Walser - ‘Tod eines Kritikers’ - hat für viel Furore gesorgt. Angefangen hat es damit, dass Walser einen Vorabdruck in der Frankfurter Allgemeinen Zeitung veröffentlichen wollte, eben jenem Blatt, in dem Marcel Reich-Ranicki den Literaturteil betreut. Es wurde mehrfach vermutet, dass Marcel Reich-Ranicki für die Hauptfigur des Romans Modell gestanden hat. Kritiker sind häufig den Agressionen der Autoren ausgesetzt, deren Werk sie in der Vergangenheit kritisiert haben.

9. Wir leben zu dritt in einer WG, zwei Mädchen und ein Junge. Eigentlich klappt das Zusammenleben ganz gut, abgesehen von der Sauberkeit. Während der männliche Mitbewohner meist seine Sachen überall liegen lässt, ist zumindest eine der weiblichen Mitbewohnerinnen äußerst ordentlich. Noch ist es zu keinem größerem Krieg gekommen, aber wir leben auch erst seit drei Monaten zusammen.

10. Es regnet schon seit einigen Tagen ohne Unterbrechung. Schön langsam fängt es an, meine Stimmung zu beeinflussen. Ich will morgens nicht aufstehen und fühle mich müde und unmotiviert. Wenn es dann bereits um drei Uhr nachmittags dunkel wird, würde ich am liebsten gleich wieder nach Hause gehen und mich mit einer Decke und einer Tasse Tee auf’s Sofa legen.

11. Gestern wurde die Stadtbibliothek wieder eröffnet. Sie war fast drei Wochen geschlossen, um die Regale wieder zu sortieren, neue Bücher einzuordnen und alte auszusortieren. Leider wurden auch einige Abteilungen komplett umgeräumt, so dass es

jetzt etwas länger dauert, sich zu orientieren. Oft können nicht einmal die Mitarbeiter genau sagen, wo bestimmte Bücher zu finden sind.

12. Die Digitalisierung ist in keinem Bereich aufzuhalten. An Computer haben sich mittlerweile fast alle Menschen gewöhnt und sehen auch die Vorteile. Digitalkameras ersetzen mehr und mehr herkömmliche Kameras, Videokassetten werden wohl auch früher oder später vom Markt verschwinden. Diesem Prozess kann man sich nur schwer entziehen.

13. Eben habe ich gehört, dass es auch im Bereich der Gebärdensprache verschiedene Dialekte gibt. Das ist zwar schwierig für Leute, die aus einer anderen Gegend kommen, aber es zeigt doch, wie ähnlich Lautsprache und Gebärdensprache sich verhalten. Ich finde es hervorragend, dass auch Gebärdensprachler durch ihren Dialekt ihre Zugehörigkeit zu einer bestimmten Gemeinschaft anzeigen können. Das ist schließlich auch eine wichtige Funktion der lautsprachlichen Kommunikation.

14. Doktoranden in Großbritannien haben ein hartes Leben. Nachdem sie mit ihrem Projekt von einem Betreuer akzeptiert wurden, müssen sie sich regelmäßig mit ihm treffen und ihre Fortschritte darlegen. Nach einem Jahr müssen sie dann eine schriftliche Ausarbeitung ihres Forschungsvorhabens vorlegen. Durch diese strikte Gliederung werden die meisten Doktoranden in drei bis vier Jahren fertig.

15. Ein Freund hat einmal behauptet, dass der Mensch eigentlich analog denkt und dass er deswegen Digitaluhren schwerer interpretieren kann. Er meinte, dass man sich bei Digitaluhren immer erst vorstellt, wie diese Uhrzeit mit Zeigern aussehen würde, um es zu verstehen. Ich weiß nicht mehr genau, woher er diese Information hatte, d.h. ob sie auf irgendwelchen wissenschaftlichen Studien beruhte oder nur Selbstbeobachtung waren.

16. Ich liebe take-away Kaffee. Es reicht schon, wenn ich auf der Straße jemanden mit diesen typischen Papp-Bechern sehe. Dann bekomme ich einen unwiderstehlichen Drang, in das nächste Cafe zu laufen und selbst einen zu kaufen. Diese Sucht ist eindeutig in Großbritannien entstanden. In Deutschland kennt man das kaum. Ich kann auch nicht genau sagen, was ich daran so besonders toll finde. Vielleicht ist es einfach das schöne Gefühl, etwas Warmes in der Hand zu halten.

17. Im Saarland wurde zum letzten Schuljahr das acht-jährige Gymnasium eingeführt. Damit erhält man das Abitur im Saarland um ein Jahr früher als in den übrigen Bundesländern. Der Vorteil dieser Reform soll sein, dass deutsche Schüler auch etwas früher mit dem Studium beginnen können. Somit wären sie international wettbewerbsfähiger.

18. Sportunterricht ist bei vielen Kindern sehr unbeliebt. Dabei treibt ein Großteil der Kinder in der Freizeit gerne Sport. Ziel des Unterrichts ist es, Freude an körperlicher Bewegung zu vermitteln. Leider erreicht der Sportunterricht oft genau das Gegenteil. Sieht man sich modernen Sportunterricht an, dann sind es wohl meist die hohen Anforderungen, die die Kinder abschrecken.

19. Nachdem es vor einigen Jahren in Deutschland kaum möglich war, einen Kindergartenplatz zu bekommen, ist die Situation heute wieder besser. Es ist schon eher der Fall, dass Kindergärten schließen, weil die Anzahl an Interessenten stark zurückgegangen ist. Paare entscheiden sich mehr und mehr dazu, kinderlos zu bleiben und sich stattdessen ihrer Karriere zu widmen.

20. Morgens ist es in der Firma immer ganz leer. Es macht Spaß, ganz allein im Büro zu sein und den Tag ohne viel Rummel zu beginnen. Im Winter kann man durchs Fenster den Sonnenaufgang beobachten. Man kann in Ruhe seine Mail beantworten, ohne von schwatzenden Kollegen abgelenkt zu werden. Die kommen meistens erst, wenn man den ersten Büro-Kaffee schon getrunken hat.

Distractor Paragraphs

1. Auf den Kanaren kann man gut wandern. Besonders attraktiv ist Teneriffa, die die größte und vielseitigste Insel ist. Wir haben einige Höhenwanderungen unternommen, die wunderschön waren. Vor allem die Aussicht war unvergesslich. Leider sind die Fotos nicht so gut geworden.

2. In Sumpfgebieten gibt es viele Libellen. Libellen gehören zu den größten lebenden Insekten. Viele Arten stellen ganz besondere Anforderungen an die Gewässer, in denen sie sich entwickeln können. Weil die Wasserqualität in vielen Fällen nicht mehr gegeben ist, sind Libellen vom Aussterben bedroht.

3. In Malaysia gibt es viel Landwirtschaft. Vor allem Reis wird in diesem tropischen Land viel angebaut. Für den Reisanbau ist ein feuchtes Klima Grundvoraussetzung. Malaysischer Reis ist in ganz Asien beliebt.

4. Klavierspielen und singen konnte Simone schon immer gut. Sie hat dafür eine besondere Begabung. Schon als sie in der Grundschule war, gab sie Konzerte. Sie hat auch keine Angst vor öffentlichen Auftritten. Ich hoffe, das ändert sich nicht, wenn sie älter wird.

5. Lämmer sind ängstliche Tiere. Selbst an ihren Hirten können sie sich nicht gewöhnen. Die Ursache könnte darin liegen, dass sie sich die Menschen nicht einprägen können

und sie daher nicht wiedererkennen. Und dabei blicken diese kleinen Schafe immer so treuherzig, dass man sie am liebsten umarmen möchte.

6. Steiniger Boden ist gut für Lilien. Dies liegt daran, dass Lilien ursprünglich in tieferen Gebirgslagen gewachsen sind. Der Boden ist meist nur mit einer dünnen Humus-Schicht bedeckt, die diesen hübschen Blumen aber genügend Halt und Nährstoffe bietet.

7. Im Urlaub fahre ich gerne nach Norwegen. Ich war das erste Mal mit meinen Eltern dort als ich gerade die Realschule abgeschlossen hatte. Mich faszinieren an Norwegen vor allem die langen Tage im Sommer und die rauhe Natur. Außerdem ist Norwegen touristisch nicht stark erschlossen. Seit damals fahre ich fast jedes Jahr hin.

8. In Italien wird man gastfreundlich aufgenommen. Es ist mir schon oft passiert, dass mich eine wildfremde Familie einfach so zum Essen eingeladen hat. Das passiert natürlich nicht in den großen Städten, weil auch in Italien inzwischen eine gewisse Hektik herrscht. Aber auf dem Dorf passiert einem so etwas noch häufiger.

9. Organisation war wichtig im Römischen Reich. Alles hatte seine Ordnung und es gab eine genau festgelegte Hierarchie. Verstöße gegen die Regeln wurden strengstens geahndet. Die Römer hatten schon früh ein gut ausgebautes Rechtssystem. Man würde es nach heutigen Maßstäben wohl nicht als gerecht bezeichnen, aber für damalige Verhältnisse war es sehr fortschrittlich.

10. Es gibt keine Marmelade ohne Zucker. Dadurch zählt Marmelade nicht unbedingt zu den gesunden Lebensmitteln. Natürlich gibt es auch Marmelade mit Zucker-Ersatz, der aber auch nicht alle Gesundheits-Richtlinien erfüllt. Marmelade bleibt also weiterhin eine süße Sünde.

11. Ende Juli beginnen in Bayern die Ferien. Dann rollt eine große Karawane Richtung Süden. Manche Bayern sind etwas geduldiger und warten ab, bis die ersten Nachbarn aus dem Urlaub zurück sind. Denn Ende August beginnt meist schon die Nachsaison. Außerdem ist es in vielen Urlaubsgegenden ruhiger, weil die Schule in allen anderen Bundesländern dann schon wieder begonnen hat.

12. Zum Einkaufen fahren viele Nord-Italiener nach Milano. Dort gibt es eine Vielzahl an erstklassigen Mode- und Schuhgeschäften. Neben günstigen Kaufhäusern findet man hier auch ganz noble Boutiquen. Da kommt jeder auf seine Kosten. Kulturell hat Mailand nicht viel zu bieten, aber man hätte ja sowieso kein Geld mehr für Theater oder Museen.

13. Der kälteste Monat ist der Januar. Obwohl die Tage schon wieder länger werden, steigen die Temperaturen nur selten über Null Grad. Eisige Winde lassen die Menschen noch mehr frieren. Komischerweise stellen gerade im Januar viele Modegeschäfte schon auf die Frühlingskollektion um. Während man frierend durch die Fußgängerzone hastet, kann man im Schaufenster dünne Frühlingsjacken und bunte Miniröcke sehen.

14. Der Waffenbesitz ist in Amerika ganz normal. Aber inzwischen wird in bestimmten Bevölkerungsschichten dieses Recht auf Waffenbesitz immer kritischer unter die Lupe genommen. Dies hat vor allem mit der ansteigenden Zahl von Verletzungen zu tun, die durch Schusswaffen hervorgerufen werden. Aber es wird wohl noch Jahre dauern, bis sich diese kritische Haltung gegenüber Waffen allgemein durchsetzt.

15. Maler arbeiten viel im Freien. Sie geben Fenstern und Hausfassaden einen neuen Anstrich, um das Material vor Witterungsschäden zu schützen. Maler müssen aus diesem Grund unbedingt schwindelfrei sein, weil sie oft tagelang nur vom Gerüst aus arbeiten.

16. Die Armenier benutzen die lateinische Schrift. Weil sie seit 1988 von der Sowjetunion unabhängig sind, wird daran auch die neueste russische Reform nichts ändern. Präsident Putin hat durchgesetzt, dass in Russland nur noch kyrillisch geschrieben werden darf.

17. Die Thronfolge der Alemannen wurde von den Merowingern diktiert. Das Herzogtum der Alemannen hat sich erst im 7. Jahrhundert unter Herzog Gottfried etabliert. Damit wurde die Kontrolle der Merowinger über die Alemannen geringer. Nach dem Tod Gottfrieds kam es zum Konflikt zwischen den Alemannen und den Karolingern.

A.4 Statistics

variable	#	non-contr.	contr.	t	p	sig?
f_0(H%)	86	168.1 Hz	167.1 Hz	0.7	.5	
f_0(L)	86	154.9 Hz	151.9 Hz	2.5	.01	
f_0(E)	86	157.3 Hz	155.1 Hz	1.8	.08	
f_0(H*)	86	218.0 Hz	227.7 Hz	3.6	$< .001$	yes
Δf_0(fall)	86	13.2 Hz	15.2 Hz	1.3	.19	
Δf_0(fall)_st	86	1.4 st	1.8 st	1.9	.06	
Δf_0(rise)	86	63.1 Hz	75.8 Hz	5.4	$< .001$	yes
Δf_0(rise)_st	86	5.4 st	6.5 st	6.9	$< .001$	yes
Δf(Erise)	86	60.7 Hz	72.6 Hz	5.2	$< .001$	yes
Δf(Erise)_st	86	5.4 st	6.5 st	6.3	$< .001$	yes
Δt(V0)	86	96.4 ms	102.2 ms	3.0	.004	?
Δt(syll)	86	152.7 ms	161.0 ms	3.0	.003	?
Δt(const)	86	436.6 ms	457.1 ms	4.3	$< .001$	yes
Δt(utt)	86	1611.6 ms	1622.4 ms	1.0	.32	
Δt(fall)	86	115.2 ms	125.5 ms	1.6	.11	
Δt(rise)	86	178.8 ms	199.3 ms	3.8	$< .001$	yes
Δt(Erise)	86	156.2 ms	173.1 ms	3.6	.001	yes
Δt(H*,V0)	86	169.5 ms	195.3 ms	6.4	$< .001$	yes
Δt(H*,V0end)	86	73.0 ms	93.1 ms	5.3	$< .001$	yes
Δt(H*,V1)	70	20.9 ms	42.4 ms	5.4	$< .001$	yes
Δt(L,C0)	86	47.0 ms	54.9 ms	1.4	.16	
Δt(L,V0)	86	-9.3 ms	-4.0 ms	1.0	.30	
Δt(L,V0end)	86	-105.7 ms	-106.2 ms	0.1	.93	
Δt(L,C1)	70	-105.5 ms	-103.6 ms	0.3	.75	
Δt(E,V0)	86	13.3 ms	22.1 ms	2.3	.06	
Δt(E,C1)	70	-81.8 ms	-75.3 ms	1.3	.19	
Δt(E,V0end)	86	-83.1 ms	-80.1 ms	1.3	.51	
slope(rise)	86	0.35	0.38	2.4	.02	
slope(Erise)	86	0.38	0.42	2.9	.005	?

Table A.2: Results of paired t-test for all variables. Significance level p=0.002.

Appendix B

Material for Perception Experiments

B.1 Details for Re-Synthesis

Test utt.	rise-version	criteria for position of L	criteria for position of H
“Armenien”	nc	mid of stressed syll.	start of foll. unstressed vowel
	early-high	””	””
	late-high	””	end of word
	late-low	end of stressed syll.	””
“Marlene”	nc	mid of stressed syll.	start of foll. unstressed vowel
	early-high	””	””
	late-high	””	begin of first vowel of foll. word
	late-low	end of word	””
“Milano”	nc	mid of stressed syll.	start of foll. unstressed vowel
	early-high	””	””
	late-high	””	begin of first vowel of foll. word
	late-low	start of foll. unstressed syll.	””
“Norweger”	nc	mid of stressed syll.	start of foll. unstressed vowel
	early-high	””	””
	late-high	””	last unstressed syll. of target word
	late-low	end of stressed syll.	””
“Römer”	nc	mid of stressed syll.	start of foll. unstressed vowel
	early-high	””	””
	late-high	””	begin of first vowel of foll. word
	late-low	start of foll. unstressed syll.	””

Table B.1: Position of L and H in the resynthesised versions of control study (rating study for all stimuli

B.2 Stimuli and Contexts for Rating Studies

B.2.1 Parallel Constructions

The parallel constructions which are used for defining a contrastive context are listed below (examples 1 to 4 are training utterances, the remaining five are target utterances):

1. Schreiner sind viel in Innenräumen beschäftigt, aber ... Maurer arbeiten viel im Freien.
   'Carpenters are often employed in interiors but ... bricklayers often work outside.'

2. Fohlen sind sehr zutraulich, aber ... Lämmer haben Angst vor Menschen.
   'Fillies are very trusting, but ... lambs are afraid of humans.'

3. In Pakistan spielt die Landwirtschaft eine große Rolle, aber ... in Indien konzentriert man sich auf Technologie.
   'In Pakistan, agriculture plays a major role, but ... India concentrates on technology.'

4. Rosen muß man regelmäßig düngen, aber ... Lilien sind sehr anspruchslos.
   'Roses need to be regularly dung but ... lillies are rather undemanding.'

5. Die Georgier haben eine eigene Schrift, aber ... in Armenien schreibt man lateinisch.
   The Georgians have their own alphabet but ... in Armenia, Latin is used.'

6. Anna liebt Ferien am Meer, aber ... Marlene fährt gern in die Berge.
   'Anna loves the sea but ... Marlene likes to go to the mountains.'

7. Pisa ist stinklangweilig, aber ... in Milano kann man gut einkaufen.
   Pisa is rather boring but ... in Milano, there are great shopping opportunities.‘

8. Finnen leben sehr zurückgezogen, aber ... Norweger geben gern Parties.
   'Finns like to stay on their own but ... the Norwegians love to have parties.‘

9. Die Germanen waren ein wilder Haufen, aber ... die Römer waren sehr organisiert.
   'The Germanics were chaotic but ... the Romans were very organised.'

B.2.2 Focus-inducing Questions for Non-Contrastive Context

Context-setting question	Stimuli
[$_\rho$ Was GEFÄLLT dir denn so am] [$_\Theta$ Maurer-Beruf]?	Maurer sind viel im Freien.
[$_\rho$ Was FASZINIERT dich denn so] [$_\Theta$ an Indien]?	In Indien konzentriert man sich auf Technologie.
[$_\rho$ Warum MAGST du] [$_\Theta$ Lilien so gerne]?	Lilien sind sehr anspruchslos.
[$_\rho$ Warum KANN man] [$_\Theta$ Lämmer nicht streicheln]?	Die Lämmer haben Angst vor Menschen.
[$_\rho$ Warum WAREN] [$_\Theta$ die Römer so erfolgreich]?	Die Römer waren sehr organisiert.
[$_\rho$ Was FINDEST du] [$_\Theta$ an Norwegern so toll]?	Die Norweger geben gern Parties.
[$_\rho$ Was MACHT [$_\Theta$ Marlene denn gerne]?	Marlene fährt gern in die Berge.
[$_\rho$ Warum FÄHRST du denn so gern] [$_\Theta$ nach Milano]?	In Milano kann man gut einkaufen.
[$_\rho$ Welches ALPHABET haben denn] [$_\Theta$ die Armenier]?	In Armenien schreibt man lateinisch.

Table B.2: Context and corresponding answers for the non-contrastive condition, small capitals indicate a high pitch accent (H*). Thematic parts are labelled with [$_\theta$ . . .], rhematic, new parts with [$_\rho$ . . .]. The four training examples are shown in the upper part, the target examples in lower part. For an English translation of the stimuli see table 4.1

B.2.3 Extended Non-Contrastive Contexts

1. Mein Praktikum werd ich dieses Jahr wohl wieder in Indien machen. Nachdem ich das ganze letzte Jahr neben dem Studium noch einen Sprachkurs gemacht habe, wird das Arbeiten in der Firma wohl noch mehr Spaß machen.
   Q: *Was* FASZINIERT *dich denn so an Indien? – In Indien konzentriert man sich auf Technologie.*

2. Ich bin richtig müde heute. Wegen der engen Termine muss man als Maurer leider oft auch am Wochenende ran. Gott-sei-Dank wird der Bau nächste Woche übergeben.
   Q: *Was* GEFÄLLT *dir denn so am Maurerberuf? – Maurer arbeiten viel im Freien.*

3. Gestern hab ich den ganzen im Garten gearbeitet. Du musst unbedingt mal vorbeikommen; mein Vorgarten ist jetzt voller Lilien.
   Q: *Warum* MAGST *du Lilien so gerne? – Lilien sind sehr anspruchslos.*

4. Gestern waren wir mit der Grundschule auf dem Bauernhof. Die Schafe hatten gerade Nachwuchs bekommen. Die Kinder waren ganz fasziniert von den Lämmern, konnten sie aber leider nicht streicheln.
   Q: *Warum* KANN *man Lämmer nicht streicheln? – Lämmer haben Angst vor Menschen.*

5. Ich such doch einen Nachmieter für das freie Zimmer in meiner WG, aber bisher haben sich nur zwei Norweger gemeldet. Mit Norwegern will ich aber irgendwie nicht zusammenwohnen.
   Q: *Was* HAST *du denn gegen Norweger? – Norweger geben gern Parties.*

6. Das Römische Reich beherrschte in seiner Blütezeit fast ganz Europa. Mühelos schlugen sie ihre Schlachten und besetzten neue Ländereien.
   Q: *Wieso* WAREN *die Römer so erfolgreich? – Die Römer waren sehr organisiert.*

7. Letzte Woche war ich wieder in Milano. Was für ein Verkehr! Beim Reinfahren stand ich über eine Stunde im Stau und das in dieser Hitze. Und dann musste ich noch ewig nach einem Parkplatz suchen.
   Q: *Warum* FÄHRST *du so oft nach Milano? – In Milano kann man gut einkaufen.*

8. Ich überleg schon die ganze Woche nach einem Geburtstagsgeschenk für Marlene. Aber irgendwie will mir gar nichts einfallen.
   Q: *Was* MACHT *Marlene denn gerne? – Marlene fährt gern in die Berge.*

9. Armenien ist ein kleines Land im Kaukasus, wo nur ca. 3 Millionen Menschen leben. Letztes Jahr waren Sie Schwerpunktland auf der Leipziger Buchmesse.
   Q: *Welches* ALPHABET *haben denn die Armenier? – In Armenien schreibt man lateinisch.*