The prosodic expression of Focus in typologically unrelated languages

vorgelegt von

Frank Kügler

aus Lübeck

zur Erlangung der Venia legendi

an der Humanwissenschaftlichen Fakultät

der Universität Potsdam

Juni 2011
Preface

The present collection of articles, submitted as a *kumulative Habilitationsschrift* at the University of Potsdam, is part of my work at the Department of Linguistics at Potsdam University including work within the DFG projects (FE 292/5-1; KU 2323/1-2; SFB 632: Project D5). The individual articles present original empirical data on several phenomena relating to the prosodic expression of focus. The main aim of these articles is to understand the rich variation languages may use, in order to express focus prosodically. These studies thus contribute to our understanding of prosodic typology, a field of research that is still growing.

I am particularly grateful to Caroline Féry for her support over the years, her inspiring ideas, and all our discussions of prosody, information structure, and its relation to linguistic theory. I also am indebted to all my co-authors for their engaged cooperation and many discussions we had during the individual experiments. In particular I would like to thank Stavros Skopeteas who introduced me to the fascinating world of Yucatec Maya, Susanne Genzel who never tires of discussing the phonetics and phonology of tone and intonation, and Anja Gollrad who is always eager to process any kind of intonational contrast. Furthermore, I am very grateful to all my colleagues at the Department of Linguistics and within the research centre 632 “Information Structure” for the inspiring exchange of ideas during these last few years. I also want to thank Lisa Pieplow-Stagg for smoothing my GermEnglish, and Stavros Skopeteas and Daniel Hole for commenting on an earlier version of this introduction.

This work is dedicated to Gösta Bruce who was a great source of inspiration and who followed many steps of this *Habilitationsschrift*, but so sadly could not experience it in the end.

Zum Schluss möchte ich mein wärmstes Dankeschön meinen drei Lieben zu Hause aussprechen, Astrid, Liv und Svea haben mich die ganzen Jahre über unterstützt, und vor allem haben sie mir immer wieder die Augen für die nicht-linguistischen Aspekte der Welt geöffnet!

Potsdam, im Juni 2011

Frank Kügler
Table of Contents

Introduction

Section I – German

Article 1: Pitch accent scaling on given, new and focused constituents in German
Caroline Féry and Frank Kügler (2008)

Article 2: The role of duration as a phonetic correlate of focus
Frank Kügler (2008)

Article 3: Production and perception of contrast in German
Frank Kügler and Anja Gollrad (accepted)

Section II – Hindi

Article 4: Focus, word order, and intonation in Hindi

Article 5: The prosodic expression of contrast in Hindi
Susanne Genzel and Frank Kügler (2010)

Section III – Yucatec Maya

Article 6: Interaction of Lexical Tone and Information Structure in Yucatec Maya
Frank Kügler and Stavros Skopeteas (2006)

Article 7: On the universality of prosodic reflexes of contrast: The case of Yucatec Maya
Frank Kügler and Stavros Skopeteas (2007)
Article 8: Encoding Information Structure in Yucatec Maya: On the Interplay of Prosody and Syntax
Kügler, Frank, Skopeteas, Stavros and Verhoeven, Elisabeth (2007)

Section IV – Akan

Article 9: On the prosodic expression of pragmatic prominence – The case of pitch register lowering in Akan
Frank Kügler and Susanne Genzel (accepted)

Article 10: Phonetic Realization of Automatic (Downdrift) and Non-automatic Downstep in Akan
Susanne Genzel and Frank Kügler (accepted)

Article 11: How to elicit semi-spontaneous focus realizations with specific tonal patterns
Susanne Genzel and Frank Kügler (2010)

Section V: A typological overview

Article 12: Introduction: Tone and intonation from a typological perspective
Sabine Zerbian, Laura Downing and Frank Kügler (2009)
Introduction

1. Preliminaries

Prosodic typology has been a growing research area in the last decade or so (e.g. Jun, 2005). However, compared to other linguistic domains in typological research, there are still too few studies to capture the broad range of variation that language employs, from a prosodic point of view. For typological comparisons a number of languages are needed that differ genetically, geographically, and are typologically diverse (e.g. Comrie, 1981; Whaley, 1997; Maddieson, 1984). This Habilitationsschrift will add data from a total of four typologically unrelated languages, three of them are less well-studied from a prosodic point of view. The body of articles of this collection will contribute to our understanding of prosodic typology comparing the prosodic realization of post-lexical prominence, in particular focus, in these four typologically unrelated languages.

There are a number of reasons why the field of prosodic typology has not reached the expertise of other fields like word order typology for instance; there is however an active research community building on the theoretical and technical advents of the last decades such that research into the prosody of individual languages rapidly and profoundly develops.

First and foremost, research in prosody is still a relatively young enterprise compared to other linguistic areas such as segmental phonology, morphology, syntax and semantics. Groundbreaking theoretical work started in the late 1970th with the autosegmental phonological theory (Goldsmith, 1976), and its application to the phonology of tone and intonation (Liberman, 1975; Bruce, 1977; Pierrehumbert, 1980). Autosegmental-metrical theory is widespread nowadays and most influential in prosodic research (cf. Gussenhoven, 2004; Jun, 2005; Ladd, 1996, 2008), and is the cornerstone of the work of the present collection of articles.
Second, yet not less crucial, for many researchers prosody still appears to be hard to analyse reliably since its phonetic correlates do carry more information than only linguistic ones (cf. the ‘half-tamed savage’ metaphor of Bolinger, 1978:475; cf. also Gussenhoven, 2004:49ff). Fundamental frequency (F0) measured as the number of vocal cord vibration per time interval (frequency in Hertz) comprises linguistic, para-linguistic, and non-linguistic information. The latter one refers to gender differences; on average, a male voice is about 100 to 140 Hz lower than a female voice. Para-linguistic information carried by F0 refers to, for instance, the emotional state of a speaker or attitudes. In addition to non-linguistic and para-linguistic information F0 conveys linguistic information such as phonological structure in terms of tone targets, which are phonetically expressed by means of F0 minima and F0 maxima over the course of an utterance, and morphological structure in terms of pitch accents that contribute to the meaning of intonation (cf. Pierrehumbert & Hirschberg, 1990; Gussenhoven, 1984). The autosegmental-metrical theory of intonation today provides a solid understanding of the linguistic parts of prosody thus disentangling the various aspects of prosody. Additionally, numerous studies on the prosody of less-well studied languages have recently been published (cf. the thematic section ‘speech prosody’ at the last ICPPhS 2007 in Saarbrücken “received by far the most submissions” compared to the total number of papers, Trouvain and Barry, 2007: vi). For less well-studied languages there is also a guide on the market on how to carry out linguistic fieldwork related to speech prosody (Himmelmann, 2006; Himmelmann & Ladd, 2008).

Third, technical constraints prevented large-scale prosodic research. It has not been until the last decade that the technical evolution facilitates prosody research tremendously. Recordings, gaining a proper speech signal for analysis, can be done on an average laptop (Ladefoged, 2003). Distinguished signal processing software is available free of charge, for instance the widely-used Praat program (Boersma & Weenink, 2011), which even allows for preparing and conducting classical perception experiments. Consequently recent research tools are easily accessible to a wide range of
researchers, and the tools are also easy to use in the field to study less well-studied languages. These factors result in an increasing body of prosodic research.

Fourth, despite the technical evolution, analysing a classical typological survey of a great number of languages experimentally as in prosody research is still almost precluded. To apply a well-designed experiment to elicit particular prosodic aspects for quantitative evaluation is both time consuming and also requires a thorough understanding of the phonology and syntax of each language. There is no speech corpus yet that allows for a controlled and quantitative prosodic analysis of a large typological language sample. For instance, controlled experimental studies in African tone languages are rare (for an overview see Zerbian, Genzel & Kügler, 2010). In addition, the limited accessibility of language resources of less well-studied languages sets constraints on prosodic typological research.

Fifth, and finally, what makes the present collection of articles interesting is the fact that it compares typologically different languages, in particular tone and non-tone languages. Simultaneous research on tone and non-tone languages is rare (Lindau, 1986; Xu & Xu, 2005; Prom-on, S., Xu, Y. & Thipakorn, B., 2009). From a typological point of view any language is assumed to have intonation (Hockett, 1963; Bolinger, 1962; Gussenhoven, 2004; but see Hyman & Monaka, 2008). Hence tone languages do not only use F0, the phonetic correlate of tone, to distinguish lexical meaning and express grammatical relations but also use F0 to express sentence-level functions and meanings. Given this, it is worthwhile to find out to what extent the phonetic cue F0 may also express post-lexical meanings. The general assumption is based on the fact that lexical and post-lexical tones share certain characteristics, which article 12 expresses as one major point of discussion in the field. One central achievement of the articles of this Habilitationsschrift is the fact that the studies are based on identical data elicitation techniques in four unrelated languages including tone languages. Hence comparison across the languages renders it possible.

The present collection of articles will contribute to prosodic typology in two major ways:
(i) The studies present an analysis of the prosodic expression of pragmatic prominence, in other words focus. In all four languages analysed, the conclusion is, that if focus is expressed prosodically, it has an effect on the pitch register either by pitch register raising (German: article 1 for production; article 3 for perception), pitch register span expansion (Hindi: article 5) and post-focal pitch register span compression (Hindi: article 4), or by pitch register lowering (Akan: article 9). Yucatec Maya is a case where no prosodic realization of focus, in terms of pitch register, occurs (articles 6, 7 and 8); speakers prefer to realise focus by means of syntactic and morphological marking (article 8).

(ii) The collection of articles provides data on prosodically less well-studied languages (Hindi, Yucatec Maya and Akan), which are analysed in an identical experimental design, which allows for a comparison with the well-studied German intonation. Typologically speaking, the data are from genetically and geographically unrelated languages; a fact that adds to the typological diversity. This Habilitationsschrift is not a classical typological study surveying a broad range of languages. It rather contributes to the growing body of languages analysed experimentally for the prosodic expression of focus.

Analysing typologically unrelated languages with a similar methodological design across languages is particularly challenging for the study of prosodic typology since comparable data provide insights in the spectrum of mechanisms that languages employ to express information structure prosodically. The present article collection contributes thus to the research that aims to develop quantitatively based, empirical prosodic descriptions. The need for quantitatively based experimental studies stems from the fact that studies based on one or two informants do not control for any speaker idiosyncrasies. General conclusions about the prosodic system of a language need to be experimentally and quantitatively based.
The issue of sentence-level pragmatic meaning is a core field of intonation research (cf. the definition of intonation provided by Ladd, 1996:6). Comparing the different strategies that the languages of the present article collection apply to express focus prosodically reveals two central claims. One, information structure does not directly affect the phonological tone structure of a language, but its effects are mediated through the pitch register. Two, the expression of focus is rather driven by communicative strategies than by phonetic universals as proposed by Gussenhoven (2002; 2004).

In the remainder of this introduction I will first introduce the concept of information structure in section 2, which the present articles are based on in order to understand the sentence-level pragmatic meanings in the sense of Ladd (1996; 2008). In section 3 a classification of the four different languages investigated will follow in terms of their typological profile as well as how the languages employ focus syntactically. Section 4 will explore the prosodic profile of the four languages and concludes with providing a taxonomy of languages in relation to the expression of focus. Section 5 will explore the relation between phonetic universals of intonation and the expression of focus concluding that speakers deviate from a neutral voice to express focus and thus attract the interlocutor’s attention. Section 6 finally argues that information structure is mediated through the pitch register, which leaves the phonological tone structure of an utterance unaffected.

2. Information structure, focus and prominence

All except three of the articles (articles 10, 11 and 12) in this collection deal with the prosodic expression of focus. Focus is here understood as a pragmatic prominence highlighting relevant information within an utterance (e.g. Chafe, 1974). In this sense, the most important information is encoded in focus. A focus of a sentence represents a word or constituent that receives prominence either by means of syntax, morphology, prosody or a combination thereof. In intonation languages such as German or English speakers dominantly use prosodic means to highlight information. This is known as
prosodic focus (Ladd, 1980; Gussenhoven, 1984; Ladd, 1996). In terms of semantics, a focus defines a set of alternatives from which one element is chosen by the speaker (Rooth, 1985, 1992; Krifka, 2008; Zimmermann, 2008). If a contrast of information is expressed, Zimmermann (2008) argues for a treatment of focus that involves additional discourse semantic notions such as hearer expectations.

A focus may appear in different positions in an utterance. If the focused constituent appears in its canonical base position one speaks of *in situ* focus. If a focused constituent appears elsewhere, hence not in its canonical base position, this is referred to *ex situ* focus. The languages of the present collection employ different *ex situ* focus strategies, which will be outlined in the next section.

To be more precise with the concept of focus we need to differentiate between different focus types. The articles of this collection follow the distinction between ‘informational’ and ‘corrective’ focus put forward by Krifka (2008). The former focus type is instantiated in answers to normal wh-questions as in (1). The analysis according to Krifka is that the question constitutes a set of propositions, and the answer is one explicit choice of the set of proposition yielding a question-answer congruency. Subscripted ‘F’ stands for focus.

(1) Informational focus

A: Who stole the cookie?
B: [PEter]$_F$ stole the cookie.     (Krifka, 2008)

Corrective focus on the other hand requires an antecedent in the previous discourse that the focus of the sentence would correct, hence restricting possible alternatives as in (2). The analysis in case of corrective focus is according to Krifka that the proposition of a previous context belongs to the set of focus alternatives. The correction then arises if the proposition differs from the context proposition, hence excluding the context proposition from the set of possible alternatives.
(2) Corrective focus

A: Mary stole the cookie.
B: (No,) [PETer] stole the cookie! (Krifka, 2008)

A focus may highlight given or new information; for a clear case of given information in focus see studies on second occurrence focus (Beaver et al., 2007; Féry & Ishihara, 2009; Baumann, Mücke & Becker, 2010). The studies of this Habilitationsschrift exclusively refer to new information in focus. If focus highlights new information, parts of a sentence may contain already old information which is usually referred to as given information, defined here as previously mentioned in the discourse (Allerton, 1978; Baumann, 2006).

The area following the focus in an utterance, i.e. given material, has specific prosodic properties. This post-focal area is under consideration in some of the articles of the present collection (articles 1, 2, 4, and 9). In these studies, any element in post-focal position is given since it was previously mentioned in the discourse. Languages differ in their post-focal prosodic expression (cf. Cruttenden, 2006 for an overview, in particular Indo-European intonation languages). Post-focal givenness may be accompanied by deaccentuation as in German for instance (article 1) or pitch register compression (Xu, 1999 for Mandarin Chinese; article 4 for Hindi; article 9 for Akan).

The next section will explore the typological profile of the languages investigated and will show how these languages employ focus syntactically before we turn to the prosodic expression of focus in section 4.

3. The languages investigated

The articles collected in this Habilitationsschrift comprise experimental studies on four typologically unrelated languages: German, Hindi, Yucatec Maya, and Akan. These languages differ from each other in a number of features. From a genetic point of view, the four languages belong to three different language families, and two of them, German
and Hindi, to the very distant genera of the same family (cf. Comrie, 1987; Haspelmath & Bibiko, 2005). From a geographical point of view, the object languages are maximally remote, and spread over four continents. And from a prosodic point of view, two of the languages, Yucatec Maya and Akan, are tone languages while the other two are not. The fact that the four languages are typologically unrelated meets a major criterion for the selection process of a language sample for typological research (cf. Comrie, 1981; Whaley, 1997; Maddieson, 1984). One language of a language family, or sub-branch of a language family, functions as a representative. Hence, the fact that the four languages under investigation are unrelated is desirable from a typological point of view.

3.1 German

German belongs to the Indo-European language family as a member of the West Germanic language group, a branch of the Germanic languages within the Indo-European language phylum (Hawkins, 1987). As a native language German is spoken by approximately 90 million speakers (Lewis, 2009), mainly in Germany, Austria and Switzerland.

It is debated whether German can be considered as an SOV language (Bach, 1962) or SVO language (Erdmann, 1990). However, German is characterized as a V2-language in terms of word order typology. German has an obligatory V2-constraint, which means that the inflected verb moves to the second position of a matrix clause (Haider & Prinzhorn, 1986). The position preceding the Verb has to be filled overtly by some constituent. A focused constituent may or may not appear in this position. Fanselow and Lenertová (2011) argue that constituents move to this position independent of information structural reasons. Focus in German is accompanied with pitch accent (cf. e.g. Uhmann, 1991; Féry, 1993), and pitch-accented constituents may appear in situ (3a), or ex situ (3b), in which case the constituent appearing in the position preceding the verb is focused; capitalization refers to the syllable that carries an accent.
(3) a. Wir haben eine Lawine gesehen!
we have an avalanche seen
b. [Eine Lawine], haben wir t gesehen!
an avalanche have we seen
‘We saw an avalanche.’

(Fanselow & Lenertová, 2011:172)

The intonation system of German is well-studied. As early as 1927, Hermann Klinghardt published the first model that covered the linguistics of German intonation (Klinghardt, 1927). Numerous further studies provided detailed phonological analyses of the German intonation system (e.g. von Essen, 1964; Wunderlich, 1988; Uhmann, 1991; Féry, 1993; Grabe, 1998; Grice, Baumann & Benzmüller, 2005; Peters, 2005), or its varieties (Barker, 2002; Gilles, 2005; Ulbrich, 2005; Peters, 2006; Kügler, 2007; Bergmann, 2008), and an even larger number of studies deal with particular phonetic or phonological aspects of German intonation (to name just a few, e.g. Atterer & Ladd, 2004; Baumann et al., 2006; Grice et al., 2009; Braun, 2005; Baumann, 2006, Niebuhr, 2007; Baumann et al., 2010; Féry & Ishihara, 2009).

The studies of the present collection vary the syntactic positions of target words carrying an accent along the possibilities outlined in (3). The articles investigate the role of F0 and duration as phonetic cues to the expression of information structure. Article 1 disentangles the influence of information structure and the role that individual tones play in the make-up of the prosodic structure of an utterance, and concludes that in German word order takes no influence on the prosodic expression of focus. Article 2 examines in detail the interaction of phonetic parameters such as the position of a word in an utterance, sentence length, and the pragmatic influence of information structure. Article 3 is concerned with the phonological structure in relation to accents under contrastive focus contexts. The higher pitch register (article 1) and longer durations (article 2) as cues for focus in speech production are also perceived (article 3) by listeners of German.
3.2 Hindi

Hindi as a member of the Indo-European language family belongs to the Indo-Iranian branch (Kachru, 1987). It is natively spoken by approximately 366 million speakers (Lewis, 2009), mainly in the central and northern part of India, but also in Bangladesh, Nepal, the United Kingdom and many other countries. In addition, Hindi is also used as a second language or a lingua franca by many Indians in non-Hindi speaking regions.

Hindi is a head-final SOV language, with relatively free word order. Constituents may be scrambled to express different information structural configurations, or for stylistic reasons. In Hindi, a focused constituent typically occupies the immediately preverbal position (4b), where wh-markers also tend to occur (Kidwai, 2000:116). According to Butt and King (1996), *in situ* focusing of a phrase in Hindi is possible with multiple foci and results in contrastive focus readings. Kidwai (2000:114ff) presents detailed arguments that focus is responsible for scrambling operations such as preposing (as XP adjunction operations) (4c).

(4) Subject focus in Hindi with the subject in the designated pre-verbal focus position in (4b), and pre-posed in (4c)
   a. kisa ne davaaii ko khariidaa?
      who ERG medicine ACC buy.PAST?
      ‘Who bought (the) medicine?’
   b. davaaii ko [graahak ne]F khariidaa
      medicine ACC customer ERG buy.PAST
   c. [graahak ne]F davaaii ko khariidaa
      customer ERG medicine ACC buy.PAST
      ‘(The) customer bought (the) medicine.’ (article 4:56)

The prosody of Hindi has been addressed by only a few studies in a non-quantitative way (Moore, 1965; Harnsberger, 1996, 1999; Harnsberger & Judge, 1996).
The two articles in the present collection (articles 4 and 5) are the first examining the prosodic expression of focus in Hindi in an experimental setting with quantitative data analysis. Focus is expressed by pitch span expansion (article 5) and post-focal pitch register compression (article 4). Together with other recent studies on the intonation of Hindi, which were conducted within our joint research project on Prosody in Parsing (Féry & Kentner, 2010; Féry and Schubö, 2010), the articles show that Hindi intonation is characterised by downstep, which is neither interrupted by focus (articles 4 and 5) nor by syntactic embedding (Féry and Schubö, 2010), and that focus causes a significant post-focal pitch register compression (article 4). Similar to German, article 4 concludes that word order variation in Hindi as presented in (4) does not affect the prosodic expression of focus. And article 5 shows that if a focused constituent cannot be syntactically moved into the designated pre-verbal focus position in case of adjective focus the pitch span related to that constituent is expanded under focus. The studies thus contribute significantly to the understanding of Hindi intonation.

3.3 Yucatec Maya

Yucatec Maya belongs to the Yucatecan branch of the Mayan languages, and is spoken in the Yucatecan Peninsula, in the states of Yucatán, Quintana Roo, and Campeche, as well as in Belize. Among the contemporary Mayan languages, Yucatec Maya is spoken by the largest population (700,000 speakers according to the 1990 census, Lewis 2009). Although the Mayan languages are extremely well documented and the most studied languages in the Americas (Campbell, 1997:165) the prosody of Yucatec Maya has only scarcely been studied.

Yucatec Maya is a head marking VOS language (Durbin & Ojeda, 1978; Skopeteas & Verhoeven, 2005), see (5a). Topicalisation and focusing are indicated by movement to designated topic and focus positions, respectively. Focus assignment is expressed by the displacement of an argument in the preverbal domain (5b). Such argument focus constructions with pre-verbal focus are analysed as cleft constructions (Bricker, 1979).
Yucatec Maya is a two-tone language (Pike, 1946; Fischer, 1976; Straight, 1976; Lehmann, 1990), and besides the articles of the present collection sentence level prosody has only been investigated by Gussenhoven & Teeuw (2008). Articles 6, 7 and 8 of this collection deal with the realisation of tone and focus in a sentence perspective. Article 6 investigates the effect of intonation on the phonetic realisation of Low and High tones. Article 7 examines the prosodic expression of focus in a context where neither syntactic nor morphological focus marking exists. In case of constituent internal focus contrast (Adj N) no movement of the focused adjective into the pre-verbal position occurs. The result is that even in this context no prosodic correlates of focus marking are expressed. The conclusion is that Yucatec Maya is ignorant to prosodic focus marking (cf. also Gussenhoven & Teeuw, 2008, for similar results). Article 8 experimentally shows that focus is realized by syntactic and morphological means rather than prosodic ones since speakers had a preference for choosing the pre-verbal position to express focus, as in (5b).

3.4 Akan

Akan belongs to the Kwa branch of the Niger-Congo family and is spoken by approximately 8.3 million people in Ghana and the Ivory Coast (Christaller, 1933; Storch, 2001; Lewis, 2009). Akan consists of several dialects, some of which are more mutually intelligible than others (Schachter & Fromkin, 1968). The dialects differ at the level of segments as well as tones (cf., Cahill, 1985; Dolphyne, 1988; Abakah, 2002, 2005; Abakah & Koranteng, 2007; among others). One of the three main dialects of
Akan is Asante Twi, which has with approximately 2.8 million speakers the majority of Akan speakers (Schachter & Fromkin, 1968, Cahill, 1985; Lewis 2009), and which our analysis is based on.

Akan is an SVO language (Boadi, 1974; Saah, 1988; Ameka, 1992; Marfo & Bodomo, 2005; Kobele & Torrence, 2006). In Akan, *ex situ* focus as in (6) is encoded syntactically by constituent fronting to the sentence initial position and morphologically by a focus marker (Boadi, 1974; Saah, 1988; Marfo & Bodomo, 2005; Ermisch, 2006; Kobele & Torrence, 2006; Amfo, 2010). In addition to focus fronting in Akan pronoun resumption in case of animate focused elements occurs (Boadi, 1974; Saah, 1988; Ermisch, 2006; Amfo, 2010), yet optionally according to Kobele and Torrence (2006).

(6) a. Kòfí rè-bóá hénà
    Kofi PROG-help who
    ‘Who is Kofi helping?’
    (Adapted from Kobele & Torrence, 2006:162)

b. [Ámái] fnà Kòfí rè-bóá nóí
    Ama FM Kofi PROG-help 3.SG
    ‘It is Ama (that) Kofi is helping.’
    (Marfo & Bodomo, 2005:185; transcription adapted)

Whether *in situ* focus, as in (7), may be expressed in Akan was debated in the literature. Marfo and Bodomo (2005:187) explicitly state that “a constituent cannot be contrastively focused *in situ* in Akan”. The authors argue that the focus marker *nà* represents the head of the extra-sentential focus phrase and does thus not appear *in situ*. However, according to Saah (1988), and Ermisch (2006) focus may be realized *in situ* as well, yet with a less emphatic interpretation compared to an *ex situ* construction (Saah 1988). Ermisch (2006) further claims that *in situ* focus occurs in case of information focus, while *ex situ* in case of corrective focus.
Article 11 of this Habilitationsschrift is concerned with this debate in more detail. In a situation-description scenario native speakers of Akan were asked to answer simple wh-questions that elicited information and corrective focus structures. One of the results was that speakers frequently used in situ focus constructions as their answers.

Like Yucatec Maya, Akan is two-tone language, and its sentence prosody is not well-studied except for local tone rules (cf. Schachter & Fromkin, 1968; Abakah, 2002; 2005). The interaction of tone and sentence-level effects such as focus (article 9) and pitch register settings (article 10) is our contribution to the understanding of intonation in a tone language. Contrary to Yucatec Maya, Akan does use intonation to express focus, and in particular speakers employ the hitherto unknown prosodic strategy of pitch register lowering in case of corrective focus (article 9). Article 9 was based on the findings of article 11 that speakers frequently use in situ strategies to express focus. Since Akan is a tone language that is sensitive to pitch register changes, article 10 elaborates the phonetic basis of pitch register changes in relation to phonological downstep.

3.5 Summary

The common theme of the majority of the articles in the present collection is the prosodic expression of focus. Typologically unrelated languages use quite a range of different prosodic strategies of expressing focus; however, the languages have in common that the phonological tonal structure is not directly affected by focus
prominence, but the effects happen at the level of the pitch register (cf. section 6 below). The articles on Hindi, Yucatec Maya, and Akan provide detailed insights into the intonation of these languages in relation to focus. The data is based on well-designed production studies, and present thus an empirical and quantitatively testable well-grounded basis. In comparison to German, the languages show that the pitch register is affected by information structure, and that the assumption of phonetic universals to be the factor of explanation does not hold. The next section will present the prosodic profile of the four languages and will provide a taxonomy for the grouping of languages according to their prosodic characteristics with respect to the expression of focus.

4. The prosodic profile of the four languages in terms of prosodic typology

From the perspective of prosodic typology numerous proposals have been made to classify languages. The classical division into tone and stress languages (non-tone or intonation languages) has been disputed for different reasons. Hyman (2006) argues that a strict typological classification follows certain constraints, that however not all languages adhere. According to the classical approach each language belongs to one and only one typological system, and there is no overlap between them; in other words a language is either a tone or a stress language. A language that does not follow this strict division is Nubi (Gussenhoven, 2006). According to Hyman (2006) Nubi cannot be grouped as either a tone or a stress language. Further, languages like Swedish or Croatian show additional features such as the word accent distinction that led to a classification of pitch-accent languages. Recently, Féry (2010) proposed the term ‘phrase language’ to classify languages such as Hindi, French, or West-Greenlandic that do only show the use of phrasal tones as opposed to intonation languages like English or German that do use both pitch accents and phrasal (boundary) tones.

These classifications refer to different levels of description. The classification system in Hyman (2001; 2006) takes the word prosodic features ‘tone’ and ‘stress’, and combines them to four logic pairings (±tone and ±stress) allowing for more than two
distinct types. The proposal by Féry (2010) and the term intonation language refer to sentence level prosodic features as opposed to word-prosodic systems.

There is however no clear-cut distinction between tone and intonation languages, or any other subdivision as pitch-accent or phrase languages as is argued in article 12 of this collection. Tone languages also make use of intonation, defined as meaningful alternations in pitch across the sentence (e.g., Xu, 1999 for focus in Chinese; Downing et al., 2004 for Chichewa; Rialland, 2009 for several African languages from distinct language families; article 9). Similarly, article 1 of the present collection argues that intonational melodies show tonal behaviour in some non-tonal languages. For reasons like these, Jun (2005) argues for a more fine-grained feature set to classify languages according to how they use pitch changes, placing them along a typological continuum rather than in absolute classes like tone vs. intonation language.

These classificatory features include phonological information at the lexical as well as the post-lexical level, and divide the prosodic features according to their functions of (i) prominence marking, and (ii) phrasing (Jun, 2005). In Table 1 the four languages of this collection are grouped accordingly. At the lexical prominence level, German is a clear case of a stress language (e.g. Féry, 1993) while Yucatec Maya and Akan are clear cases of tone languages (e.g. Pike, 1946; Dolphyne, 1988). Hindi however would rather be characterized as having neither stress nor tone employing post-lexical tone features to demarcate prosodic words (Féry, 2010), although the feature ‘stress’ is disputed in Hindi (cf. article 5; Ohala, 1986; Nair, 2001). For Akan, some researchers assume stress to be present (Christaller, 1933; Purvis, 2009; Anderson, 2009). However, Table 1 expresses a ‘?’ for Akan stress since to our knowledge, no comprehensive analysis of stress in the sense of Hayes (1995) exists, and the exact phonetic details of the “emphasis put on a syllable” (Christaller, 1933:XXVIII) remain unclear.
Table 1. Prosodic features of the four languages of the present collection according to Jun (2005:444); the feature set is adapted. ‘x’ means the feature exists in that language, ‘?’ status unclear, ‘–’ not applicable, (PW) refers to a higher prosodic unit than the foot, i.e., the prosodic word.

<table>
<thead>
<tr>
<th></th>
<th>Prominence</th>
<th>Prosodic unit</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Lexical</td>
<td>Post-lexical</td>
</tr>
<tr>
<td></td>
<td>tone</td>
<td>head</td>
</tr>
<tr>
<td></td>
<td>stress</td>
<td>edge</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>German</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Hindi</td>
<td>?</td>
<td>x</td>
</tr>
<tr>
<td>Yucatec Maya</td>
<td>x</td>
<td>–</td>
</tr>
<tr>
<td>Akan</td>
<td>x</td>
<td>?</td>
</tr>
</tbody>
</table>

At the lexical prosodic unit level German and Hindi employ the foot (e.g. Wiese, 1996) respectively the prosodic word (e.g., Moore, 1965; article 4) as their domain of structure, but it is the syllable in Yucatec Maya (Pike, 1946) and Akan (Dolphyne, 1988) which is the tone bearing unit. Post-lexically, all languages have a level of intonation phrase or utterance which is not displayed in the table for reasons of redundancy. Minor phrases are attested for German, Hindi, and Yucatec Maya. The exact status of a phrasing in Akan remains unclear though. At least, the experimental findings of article 9 do not hint at durational correlates of phrasing in relation to focus.

Post-lexical prominence marking in the four languages differs, which the present research has shown. German and Akan mark the head of a prosodic unit, expressed by means of association of a pitch accent with the prominent syllable in German (among others Féry, 1993; Grice, Baumann & Benzmüller, 2005; article 1, and 3), and of a new pitch register line in Akan (article 9). In Hindi, phrase tones mark the edge of prosodic units (article 4), and the phonetic realization of tones affects the pitch register (article 5).
In Yucatec Maya, no prosodic marking of prominence is expressed (articles 6, 7 and 8; Gussenhoven & Teeuw, 2008).

Comparing the four languages certain differences in the prosodic expression of focus become obvious. Given the fact that any language is assumed to have intonation (Hockett, 1963; Bolinger, 1962; Gussenhoven, 2004), which refers to the linguistically structured and phonologically entrenched expression of sentence-level pragmatic meanings (Ladd 1996; 2008), the phonetic cue F0 may also express post-lexical meanings despite the fact that F0 distinguishes lexical meaning and expresses grammatical relations by means of tone in tone languages. In terms of sentence-level meanings, the phonetic cue F0 highlights information (focus) and partitions utterances by means of phrase and/or boundary tones. Surveying the literature we identify three groups of languages on the basis of their distinct prosodic expression of focus.

In one group of languages focus marking is accompanied by changes in F0 scaling, i.e., an enhancement of the tonal register. This group includes languages that express focus with higher or later pitch peaks such as German (see Braun, 2005; Baumann et al. 2006; article 1 and 2; Kügler & Genzel, MS), English (see Bartels & Kingston, 1994; Cooper, Eady & Mueller, 1985; Eady & Cooper 1986; Eady, Cooper, Klouda, Mueller & Lotts, 1986), or a number of tone languages such as Mandarin Chinese (Xu, 1999; Liu & Xu, 2005; Chen et al., 2009), Thai (Pan, 2007), Vietnamese (Jannedy, 2007), Hausa (Leben, Inkelas & Cobler, 1989; Inkelas & Leben, 1990), or Beaver (Schwiertz, 2009). Xu (1999) showed that all four tones in Mandarin Chinese are articulated with expanded pitch register, e.g. a H tone is raised higher, and a L tone is lowered. In addition to a more enhanced production of tones under focus the tonal register after the focus is compressed. Also, longer durations of focused words have been reported for Thai (Pan, 2007) and Vietnamese (Jannedy, 2007). As another strategy, the Curaçao dialect of Papiamentu, a Caribbean Creole with lexical tone contrasts, employs a particular focus pitch accent to mark prominence (Remijsen & van Heuven, 2005), which resembles the Swedish focal accent (Bruce, 1977).
A second group of languages employs a different strategy to express focus, which refers to prosodic domain structure, i.e. an insertion of a phrase break before or after a focused constituent. In Bengali, for instance, a boundary tone functions as demarcation of a focus phrase (Hayes & Lahiri, 1991). Other languages of this group are Pero (Frajzyngier, 1989), Tangale (Kidda, 1993), Chichewa (Kanerva, 1990; Downing, Mtenje & Pomponio-Marschall, 2004; Downing, 2008), Kammu (Karlsson, House, Svantesson & Tayain, 2007), or Beaver (Schwiertz, 2009). A phrase break is realized by means of phonetic cues such as a pause, final lengthening (Martin, 1970; Lehiste, 1972; Wightman, Shattuck-Hufnas, Ostendorf & Price, 1992), and/or F0 register resetting (e.g. the Bantu language Chichewa: Kanerva, 1990; Downing et al., 2004; Downing, 2008; and some Kwa languages spoken in Côte d'Ivoire, yet not Akan: Leben & Ahoua, 2006). In Kammu, a focus is signalled by means of a H boundary tone (Karlsson et al., 2007).

Contrary to languages of the first two groups, a third group of languages does not use prosodic means for the expression of prominence at all. These include the Bantu language Northern Sotho (Zerbian, 2006), the Mayan language Yucatec Maya (article 6, 7 and 8; Gussenhoven & Teeuw, 2008), Navajo which belongs to the Athabaskan language family (McDonough, 2002), and Wolof, a Niger-Congo intonation language with no lexical tones and no prosodic reflex of focus marking (Riallaid & Robert, 2001). Hartmann and Zimmermann (2007) present evidence from production data and perception tests that also Hausa does not use prosodic means for the encoding of focus. Their results go against the findings of Leben et al. (1989) and Inkelas and Leben (1990).

This review shows that besides tonal distinctions of lexical and/or grammatical functions tone languages do use prosodic means for the expression of post-lexical pragmatic meanings, yet not all tone languages necessarily employ prosodic means for the expression of post-lexical pragmatic meanings. At the same time a non-tone language such as Wolof does not seem to employ particular intonational strategies of focus marking. The collection of articles in this Habilitationsschrift comprises
languages from two of these groups: German, Hindi, and Akan belong to group 1 (articles 1, 2, 3, 4, 5, 9), and Yucatec Maya to group 3 (articles 6, 7, 8).

In sum, the prosodic profile of the four languages investigated has shown that the prosodic expression of focus is not uniform across the languages. In the next section we discuss the relation between phonetic universals of intonation and the expression of focus prominence.

5. Phonetic universals and prominence

Comparing the prosodic expression of focus across the four languages reveals a further aspect that relates to the universals aspects of the interpretation of pitch variation. According to Gussenhoven (2002; 2004) intonation comprises phonetic universal, and language specific phonological aspects. The phonetic universal aspects denoted as the ‘biological codes’ (Gussenhoven 2002; 2004) are those universal paralinguistic form-function relations and dimensions of meaning that refer to the production process of F0. Gussenhoven distinguishes the Frequency Code, the Production Code, and the Effort Code. The Frequency Code correlates with the size of the larynx and with the rate of vocal cord vibration. The Production Code refers to the intensity of the pitch and its decline from the beginning of a speech act to its end. The Effort Code finally refers to the amount of energy used in speech production.

One commonplace grammaticalisation of the Effort Code is the expression of focus. The effort code predicts that the more articulatory effort is used for articulation to express prominence, here focus, the higher the fundamental frequency gets. Hence, the prosodic expression of focus is predicted to yield higher F0 peaks, which indeed is the case for English (Gussenhoven, 2004), and German (e.g., articles 1, and 3). In German, for instance, focused elements are marked by prosodic prominence in terms of pitch register variation (article 1 for production, article 3 for perception), and/or duration (article 2) in order to attract the listeners’ attention to a particular information of the utterance.
The prediction however does not hold true in the other three languages under consideration. Contrary to predictions, in Akan tones are realized at a lower register. This concerns both High and Low tones (article 9). Also for Hindi, this prediction does not hold true: In article 4 of this collection there is no evidence for higher F0 in Hindi that is related to focus; article 5 more specifically states that F0 height is not necessarily increased but the pitch span between the Low and the High tone. Finally, in Yucatec Maya no prosodic reflex of focal prominence is expressed at all.

Apart from Yucatec Maya, which according to the our grouping of languages above (cf. section 4) belongs to a class of languages that do refrain from using prosodic means for the expression of focus, the data from Akan suggest a different interpretation: Given the fact that the prosodic expression of focus differs from a neutral register (wide focus) it seems plausible that it is not the direction of change, in other words F0-raising as in Germanic languages and as predicted by the effort code that does matter, but the deviation from the neutral register. Speakers seem to have a choice to either raise their voice as frequently done in English or German, or lower their voice as frequently done in Akan.

Further evidence for the deviation account comes for instance from Mandarin Chinese. Xu (1999) shows that High tones are raised, which is fully in line with Gussenhoven’s universal account; for Low tones however, the tones are lowered, which would go against the universal account. These facts show that the direction of change is not unidirectional. Languages may have different strategies to deviate from a neutral voice. And even within a language, cf. Mandarin Chinese, different tones may undergo different changes under the expression of focus. This would impressionistically also hold for German or English in case of L* pitch accents, which appear to be realised lower under focus (for English cf. Liberman & Pierrehumbert, 1984).

To sum up, the claim of this Habilitationsschrift is that the expression of prominence follows communicative rather than phonetic universal constraints such that speakers vary their voice from a neutral register in order to attract attention of their interlocutor. Furthermore, the direction of change may be determined by the
phonological information of the tone that is subject to variation: a H* pitch accent in
German may rise, a L* on the other hand may be lowered. Finally, the research on the
prosodic expression of focus prominence in the four languages has shown that the pitch
register is affected, which will be outlined in the next section.

6. Prosodic expression of focus in terms of pitch register

A major issue that the main body of articles of the present collection tackles is the
question of how lexical tones and intonational tones interact. In particular for Yucatec
Maya and Akan the question is how much intonation do these languages use apart from
their phonological and grammatical use of tone in terms of F0. Opposed to tone
languages, for German and Hindi the question is how the post-lexical tones (pitch
accents and/or boundary tones) interact with the expression of focus. The overall
assumption from the research presented in this collection is that if focus is expressed
prosodically it is mediated through the pitch register.

The term pitch register is used here following the definition proposed by Ladd
(1990). The pitch register represents abstract reference lines, relative to which local
tonal targets are scaled (cf. Clements, 1979; tonal space between black lines in Figure
1). These reference lines are located in the pitch range of a speaker, which is a speaker
specific parameter (Ladd, 1990; red lines in Figure 1). Any change of pitch range refers
to the emotional state of the speaker, while linguistic factors such as downstep (register
shift illustrated in Figure 1), or the realisation of pragmatic prominence (focus) change
the pitch register.

In the languages investigated, focus prominence does not affect the tonal structure
of an utterance. In German a typical tonal structure of an utterance with more than one
pitch accent consists of pre-nuclear rising pitch accents and a nuclear falling one (e.g.,
Féry, 1993). In case of focus, pre-focal constituents still carry a rising pitch accent, and
the last accent is a falling one, though the scaling of all the accents differ (article 1; and
article 3 for perception). Of course other accents than a falling one may be realised in
case of focus (for a quantitative distribution of different accent types see Grice et al., 2009). Article 1 of this collection shows that the scaling of High tones in German remains constant as a function of word order variation, and that focused constituents show a significantly higher scaling than corresponding accents associated with constituents in a neutral or wide focus condition. Article 2 of this collection shows that in addition to scaling differences systematic lengthening effects of focus occur on accented syllables though not linearly distributed over the segments. And article 3 of this collection shows that scaling differences in relation to focus are perceived by native listeners: in a semantic congruency paradigm subjects were asked to rate the accuracy of question-answer pairs. Congruent question-answer pairs, i.e., a question asking for corrective focus with an adequate answer including a corrected constituent, were rated significantly more often as congruent, than incongruent question-answer pairs.

**Figure 1.** Illustration of pitch range (red lines represent the top and bottom line of the pitch range), and pitch register reference lines (black lines, which refer to upper and lower register lines), adapted from Ladd (1993:119).

In Hindi the dominant prosodic pattern consists of downstep (article 4 and 5; Féry & Schubö, 2010). The downstep pattern is not interrupted regardless of any syntactic structure or pragmatic influence of focus. The pitch span is increased in case of *in situ* focus (article 5), and post-focally the pitch register is compressed, yet tonal rises are still realized. The original tonal make-up of an utterance is thus not affected.
The same holds for Yucatec Maya and Akan, as well as in other tone languages such as Mandarin Chinese, for instance (Xu, 1999). Any lexical tone maintains its phonological information in case of prominence in order to maintain lexical and/or grammatical contrasts. What can be observed is that the pitch register is affected in case of focus. In Akan a register lowering is realized, hence lexical High tones are still realized with higher F0 than lexical Low tones, yet in a lower register (article 9). A general conclusion to be drawn from these studies on the prosodic expression of prominence is that the pitch register needs its own auto-segmental tier to model the focus effects. Similar proposals have been made for the analysis of downstep (e.g. Snider, 1999). Along these lines, focus seems to affect a phonologically determined pitch register specification.

7. Conclusions

The conclusion for prosodic typology of this collection of articles is that investigating less-well studied languages provides insights into the diversity of language structures, with different theoretical implications. Comparing the prosodic expression focus applying a unified method across the languages allows for thorough comparisons of the results. From this comparison the proposal arose that the expression of focus is prosodically achieved by means of pitch register variations.

Insights from Akan compared to other languages led to the conclusion that across languages speakers may rather employ communicative strategies to highlight information in discourse than activating phonetic universals of intonation. The very fact that in Akan pitch register lowering is used to express focus goes against claims based on the biological codes (Gussenhoven, 2004). The present collection of articles presents case studies that show that the deviation from a neutral register matters for the expression of prominence, irrespective of the direction in which this deviation goes.
References


Allerton, D. J. (1978). The notion of `givenness' and its relations to presupposition and to theme. Lingua 44, 133-168.


Harnsberger, J. D. (1999). The role of metrical structure in Hindi intonation. South Asian Language Analysis Roundtable, University of Illinois, Urbana-Champaign


Kügler, F. & Genzel, S. (MS) Sentence length, position, and information structure effects on segmental duration in German. Ms, University of Potsdam.


Kügler


Appendix

Section I: German

Article 1. Pitch accent scaling on given, new and focused constituents in German
Caroline Féry and Frank Kügler 2008.

This article is joint work, and my additional contribution to this article was
the conduction of the production study, and the phonetic and statistical
analysis of the data. Together we designed the study and developed the
interpretation of the data.

Article 2. The role of duration as a phonetic correlate of focus.
Kügler, Frank 2008.
In Barbosa, P. A., Madureira, S., and Reis, C. (eds). Proceedings of the
Speech Prosody 2008 Conference. Campinas, Brazil: Editora RG/CNPq,
591-594.

Article 3. Production and perception of contrast in German.
Kügler, Frank and Gollrad, Anja, 2011.
*Accepted* for publication in the Proceedings of ICPhS 2011, Hong Kong.

My contribution to this article was the design of the study, the analysis of
the production data, the manipulation of the stimuli as well as to oversee
the analysis of the perception data. I also put the data into the relevant
background of the literature. Anja Gollrad conducted the perception study.
Section II: Hindi

Article 4. Focus, word order, and intonation in Hindi
Journal of South Asian Linguistics, 1, 53-70.

My contribution of this article was the initiation of this study, and to oversee the phonetic as well as the statistical analysis. In joint work we designed the production study. As the corresponding author of this article I had the main responsibility of this research. Umesh Patil did the recordings of the speakers. It was decided in the project, that the PhD students appear as first authors (Patil, Kennter & Gollrad), and the PI’s of the project appear after the PhD students with me being first as the corresponding author.

Article 5. The prosodic expression of contrast in Hindi
Susanne Genzel and Frank Kügler 2010.
Speech Prosody 2010, Chicago.
(http://speechprosody2010.illinois.edu/program.php#100143)

My contribution was the initiation of the design of this study, which originally was the MA thesis of Susanne Genzel. I supervised this study, discussed the relevant phonetic analyses as well as the statistics.
Section III: Yucatec Maya

Article 6. Interaction of Lexical Tone and Information Structure in Yucatec Maya
Frank Kügler and Stavros Skopeteas 2006.

My contribution to this article was the design of the study, the phonetic and statistical analyses as well as the phonological interpretation. Stavros Skopeteas elicited the data, and introduced me to general aspects of Yucatec Maya.

Article 7. On the universality of prosodic reflexes of contrast: The case of Yucatec Maya
Frank Kügler and Stavros Skopeteas 2007.

My contribution was the design of this study, phonetic and statistical analysis of the data. Together we elicited the speech data, and put the analysis into the background.

Kügler, Frank, Skopeteas, Stavros & Verhoeven, Elisabeth 2007
My contribution was the design of the prosodic part of this study, and the data for this part was elicited together with Stavros Skopeteas. Together we discussed the elicitation procedure for the focus realisation study, which Stavros Skopeteas and Elisabeth Verhoeven conducted. All of us discussed the implications of the tonal and syntactic data for the interpretation of focus realisation in Yucatec Maya.

**Section IV: Akan**

**Article 9.** On the prosodic expression of pragmatic prominence – The case of pitch register lowering in Akan
Frank Kügler and Susanne Genzel 2011.
*Accepted, Language and Speech.*

My contribution of this article was the design of the studies, the supervision of the phonetic analysis, the anchoring of the study in the relevant literature as well as the conduction of the statistical analysis. Data elicitation was done by Susanne Genzel. Together we checked the transcriptions and discussed the analysed data. I am the corresponding author.

**Article 10.** Phonetic Realization of Automatic (Downdrift) and Non-automatic Downstep in Akan
Susanne Genzel and Frank Kügler 2011.
*Accepted for publication in the Proceedings of ICPHS 2011, Hong Kong.*

My contribution to this article was the design of the study, the supervision of the analysis as well as the anchoring of the study in the literature. Data elicitation was done by Susanne Genzel. Together we discussed the outcome and proposed the analysis.
Article 11. How to elicit semi-spontaneous focus realizations with specific tonal patterns
Susanne Genzel and Frank Kügler 2010.

My contribution to this article was the initiation of the study, and the methodological decision for the task to elicit focus realisations. We hired a professional painter to draw the pictures. Susanne Genzel conducted the task. We did the final interpretation of the data and its presentation together.

Section V: A typological overview
Article 12. Introduction: Tone and intonation from a typological perspective
Sabine Zerbian, Laura Downing & Frank Kügler 2009.

This article was joined work, and my main responsibility was on the intonation part. We organized the TIE-2 conference together of which we selected some of the studies to appear in the special issue of Lingua. Sabine Zerbian was the main editor of the special issue. Together we argued for similarities between tone and intonation languages.
Section I
German
Pitch accent scaling on given, new and focused constituents in German

Caroline Féry*, Frank Kügler

Department of Linguistics, University of Potsdam, Karl-Liebknecht-Straße 24/25, D - 14476 Golm, Germany

Received 20 October 2006; received in revised form 28 April 2008; accepted 6 May 2008

Abstract

The influence of information structure on tonal scaling in German is examined experimentally. Eighteen speakers uttered a total of 2277 sentences of the same syntactic structure, but with a varying number of constituents, word order and focus-given structure. The quantified results for German support findings for other Germanic languages that the scaling of high tones, and thus the entire melodic pattern, is influenced by information structure. Narrow focus raised the high tones of pitch accents, while givenness lowered them in prenuclear position and canceled them out postnuclearly. The effects of focus and givenness are calculated against all-new sentences as a baseline, which we expected to be characterized by downstep, a significantly lower scaling of high tones as compared to declination. The results further show that information structure alone cannot account for all variations. We therefore assume that dissimilatory tonal effects play a crucial role in the tonal scaling of German. The effects consist of final f0 drop, a steep fall from a raised high tone to the bottom line of the speaker, H-raising before a low tone, and H-lowering before a raised high tone. No correlation between word order and tone scaling could be established.

1. Introduction

Researchers agree on classifying German as an intonation language (see the typologies in Gussenhoven, 2004; Jun, 2005; Ladd, 1996, among others). It has lexical stress and uses different types of postlexical pitch accents and pitch accent sequences as well as boundary tones for the expression of pragmatic contrasts (Baumann, 2006; von Essen, 1956; Féry, 1993; Mayer, 1997; Peters, 2006; Pheby, 1980). Various aspects of German dialect intonation are well studied (Barker, 2002; Gilles, 2005; Kügler, 2007; Peters, 2006), but here we restrict ourselves to Standard German as spoken in the Berlin-Brandenburg region. Thus far, the bulk of the analysis of the tonal structure of this language has been directed towards the tone inventory, the alignment of these tones with segments, and variation in the use of tones. Tone scaling across accents and across phrases has not been extensively studied, though some research has been done on the subject (see Katz & Selkirk, 2005; Ladd, 1990 and Selkirk, 2006 for English; Féry & Truckenbrodt, 2005; Kügler, Féry, & van de Vijver, 2003; Truckenbrodt, 2002 for German). A number of studies on German have specifically investigated...
downstep (Grabe, 1998; Truckenbrodt, 2002, 2004).\(^1\) To our knowledge, however, no study has been published so far that addresses tone scaling in German in relationship with information structure more than in passing, and purely tonal effects like H-raising and H-lowering have been left unnoticed. However, to fully understand the tonal structure of a language such as German, a quantification of the combined effects of syntax, information structure and of the tones themselves is necessary. This is the subject of the present paper.

Information structure is an important component of our study. The notions used here are restricted to focus and givenness, which are regulated by the context in which the experimental sentences were embedded. Focus appeared in two variants: wide (or all-new) and narrow (see Ladd, 1980 for this distinction). In an all-new sentence, no element has been mentioned in the preceding context or was especially prominent in the common ground of the protagonists. A narrow focus was induced by a context asking explicitly for one or more arguments, or for the verb. By contrast, a given constituent has been mentioned in the question or the context introducing the target sentence. A semantically informed definition of focus and givenness is beyond the scope of this paper, but we base our view of information structure on the work of Rooth (1985, 1992), Schwarzschild (1999), and the papers in Féry, Fanselow, and Krifka (2006), among others.

Default sentence accent assignment, as well as accents motivated by information structure, have been examined extensively for German (for instance by Bierwisch, 1966; Büring, 2001; Cinque, 1993; Féry & Samek-Lodovici, 2006; Fuchs, 1976; Gussenhoven, 1992; Jacobs, 1993; von Stechow & Uhmann, 1984). In an all-new sentence, accents are assigned on metrically prominent positions in the sentence, determined solely on the basis of the syntactic structure of the sentence. Every argument of a verb is accented and the verb itself may be accented or not, depending on the phrasal integration of the verb and its immediately preceding argument, an optional process (see, for instance, Fuchs, 1976; Gussenhoven, 1992; Jacobs, 1993). In the all-new pattern, all arguments are expected to be accented. According to Pierrehumbert (1980) and Liberman and Pierrehumbert (1984), scaling of tones in English takes place from left to right, downstepping a high tone relative to a preceding high tone. The calculation of the phonetic height of a tone takes the local left and right phonological context into account, as well as the left phonetic context.

At the end of this subsection, a series of hypotheses are formulated, that summarize our expectations concerning the experimental results. In German, as well, downstep of a series of accents has been taken to be the unmarked pattern (see, for instance, Féry, 1993; Truckenbrodt, 2002, 2004). The all-new sentence condition can thus function as a baseline for comparison of the effects of information structure (Hypothesis 1).

In sentences with narrow foci, main accent is assigned to the last focused constituent. Based on the findings for different Germanic languages that focus generally boosts accents (Baumann, Grice, & Steindamm, 2006; Cooper, Eady, & Mueller, 1985; Eady & Cooper, 1986; Eady, Cooper, Klouda, Mueller, & Lotts, 1986; Gronnum, 1992; Thorsen, 1985) we expect that a narrow focus raises a pitch accent in all positions, as formulated in Hypothesis 2. In the same way, as focus is expected to raise the values of pitch accents, givenness is expected to lower them (Bartels & Kingston, 1994; Brown, Currie, & Kenworthy, 1980; Cruttenden, 2006; Féry & Ishihara, 2005, to appear; Ladd, 1980, 1983), though prenuclear and postnuclear givenness have completely different effects. We expect given constituents to be realized lower in prenuclear position and to be deaccented in postnuclear position (Hypothesis 3). A prenuclear given constituent is associated with a pitch accent, even if this pitch accent is comparatively low. A postnuclear given constituent is, by contrast, realized with a low and flat contour and is analyzed as being deaccented (see also Cruttenden, 2006; Ladd, 1996; Liberman & Pierrehumbert, 1984 for English).

In Fig. 1, the expectations related to the effect of narrow focus and givenness at different places in a sentence with multiple arguments and a final verb are displayed. The dotted line shows the regular downstep in an all-new sentence. This is the pattern assumed to be the unmarked intonation according to Hypothesis 1. Fig. 1a stands for an initial narrow focus on argument one (A1). A1 and A2 and stand for Argument 1 and 2,\(^1\)

\(^1\)The reverse effect, upstep, has also been claimed to exist in German. Truckenbrodt (2002) proposes that upstep is an optional phenomenon limited to the last pitch accent or phrasal boundary line in a non-final intonation phrase. He analyzes upstep as an undoing of a downstepped sequence. In Truckenbrodt’s account, upstep must be preceded by downstepped pitch accents and followed by an intonation phrase (see also Féry & Truckenbrodt, 2005). Since the study presented here concentrates on sentences consisting of only one intonation phrase, we never find this kind of upstep in our data, although we do find a similar effect of raising of the last accent in an intonation phrase.
respectively. \( A_1 \) is higher than in the all-new configuration (because it is narrowly focused, Hypothesis 2), argument 2 \( A_2 \) and the verb \( V \) are lower (because they are given and deaccented). Fig. 1b displays the expected effect of a narrow focus on a medial argument. \( A_1 \) is expected to be lower than in the all-new configuration (because it is given and prenuclear, Hypothesis 3), \( A_2 \) is expected to be higher (it is narrowly focused), and \( V \) lower. Fig. 1c shows the changes in pitch accents when the verb is narrowly focused. The preverbal arguments are lower than in the all-new condition and downstepped relative to each other, since both are given, and there is an upstep on the verb.

In addition to the effects on \( f_0 \) we also expect an effect of information structure on duration (Hypothesis 4). Durational influences of focus have been shown for English (Beaver, Clark, Flemming, Jaeger, & Wolters, 2007; Eady & Cooper, 1986), for German (Baumann et al., 2006; Kügler, 2008) and in the case of second occurrence focus in German (Féry & Ishihara, 2005, to appear). We therefore expect a lengthening effect of focus to occur in our data as well.

The last, more speculative, hypothesis for information structure relates to word order. It is assumed that non-canonical word orders, that is those in which an argument other than the nominative is initial, are always triggered by marked information structure, and that, for this reason, scaling of pitch accents should be affected by changes in word order (see Hypothesis 5).

Following Liberman and Pierrehumbert (1984) for English and Grabe (1998) for German we predict a final constant value at the end of declarative sentences (‘low endpoint’ in the terminology of Liberman & Pierrehumbert), regardless of the length of the sentence or of its focus pattern (see Maeda (1976) for introducing this notion and Menn and Boyce (1982) for discussion, Hypothesis 6).

Finally, based on the discussion of information structure and sentence accent assignment above we expect that the combined effects of syntax and information structure explain all tone values found in the experiment (Hypothesis 7).

The hypotheses are formulated as follows:

1. All-new sentences induce an unmarked accent pattern in which all constituents carry an accent, and there is a downstepping pattern throughout; the verb is optionally accented.
2. Narrow focus on a constituent induces raising of the corresponding high tone.
3. Prefocal givenness induces lowering of the high tone and postfocal givenness induces final deaccenting.
4. A constituent with narrow focus is longer than the corresponding constituent in an all-new sentence or when it is given.
5. A changed word order increases the pitch height of the prominent constituent.
6. A final constant value (‘low endpoint’) is reached at the end of declarative sentences, regardless of the length of the sentence or of its information structure.
7. The pitch height of all individual pitch accents are unambiguously determined by default prosodic phrasing and information structure.

The next section outlines the methodology of the production experiment. Section 3 reports the results and quantifies the effects of newness, givenness and narrow focus on the scaling of high tones in pitch accents in German. The results of the production experiment revealed that information structure alone cannot account for the various patterns in our data. Therefore, we propose additional dissimilatory tonal effects that account well for the additional effects. In Section 4, a model of pitch accent scaling is proposed in terms of register reference lines that integrate the effects of syntax and information structure, as well as the purely tonal effects introduced in Section 3.
2. Method

2.1. Material

The aim of the experiment reported in this section was to investigate the scaling of pitch accents in sequences of accented and deaccented words in relationship to their information status, their place in a tone sequence and the number of realized accents. The sentences used in the experiment have a simple syntactic pattern: they are verb-final sentences introduced by a complementizer weil ‘because’, which contain one, two, or three arguments plus the verb.

The reason for choosing an embedded verb-final pattern is that this structure is considered to be the most unmarked word order of German (Lenerz, 1977). A single argument is always a nominative. When there are two arguments, they are nominative and accusative, or nominative and dative. Sentences with three arguments contain a nominative, a dative and an accusative. In the experiment, all sentences were introduced by a context, which was just a single question (in the case of all-new sentences) or an introductory sentence followed by a question. The arguments were masculine animal names, so that case was unambiguously recognizable on the article (German feminine and neuter articles have the same morphological form in nominative and accusative, but not the masculine). The five animal names used were trochaic with a final schwa syllable (Hammel ‘sheep,’ Hummer ‘lobster,’ Löwe ‘lion,’ Rammler ‘buck,’ Reiher ‘heron’). Only particle verbs were used, with detachable and stressed particles. The complex verbs were always in the participial form and were located in the sentence-final position (angefangen ‘begun,’ eingeladen ‘invited,’ nachgelaufen ‘followed,’ vorgestellt ‘introduced’). They vary in the number of arguments they require and in the cases they govern.

Examples of the sentences are given in (1)–(4), embedded in the appropriate context. The material between curly brackets is a translation of the context sentences. The target sentences are in italic; underlined constituents indicate focused material. N stands for nominative, A for accusative, D for dative, and V for verb.

(1) NDV (focus on two arguments but not the verb)
{The animals like to play ‘catch.’ One animal has to follow another one. Why were they happy this time?}
Weil der Hammel dem Rammler nachgelaufen ist.
‘Because the sheep followed the buck.’

(2) NAV (all-new condition)
{Why were the animals happy?}
Weil der Hammel den Rammler eingeladen hat.
‘Because the sheep invited the buck.’

(3) NADV (focus on the dative complement)
{The sheep wanted to introduce the buck to the lion. Why didn’t he do this?}
Weil der Hammel dem Rammler dem Hummer vorgestellt hat.
‘Because the sheep introduced the buck to the lobster.’

(4) NV (focus on the verb)
{The animals don’t like to fight. Why are they angry with the lobster?}
Weil der Hummer angefangen hat.
‘Because the lobster started (a fight).’

Three parameters, which are listed in Table 1, were systematically varied. These parameters comprise the number of arguments, word order and information structure. Altogether 26 conditions were created. Each condition was realized in five versions, in which the five animal nouns were permuted in a systematic way, but the verb remained constant. In total, this resulted in 130 sentences (26 conditions × 5 renditions) per speaker. In total 2340 sentences (130 sentences × 18 speakers) were recorded.

The missing constellations comprise very unnatural possibilities, like focus on an initial scrambled constituent.
2.2. Recordings

The entirety of the data used in this paper was collected in one experiment, run individually with 18 speakers. All were female students at the University of Potsdam, Germany. They were monolingual speakers of German in their 20s, coming from the Northern area of Germany. They were paid for their time.

The target sentences were recorded in a soundproof booth on a DAT tape recorder. A set of instructions familiarized the subjects with the process and had them practice with four examples. After the instructional part, the experimenter left the subject alone in the room. The subject went through the experiment in the form of a Powerpoint presentation in a self-paced manner. The speakers read the sentences on a screen as the answers to questions which were presented both visually and acoustically over headphones: they heard and read a question on a computer screen, pressed the return key, and read aloud a target sentence presented on the next slide. The items to be accented were underlined in order to minimize errors. The context sentences had been recorded previously. They were spoken by the second author, a trained phonetician and native speaker of Standard German in his thirties, who also comes from the northern part of Germany. He was recorded in a soundproof booth on a DAT recorder. He spoke naturally, at a conversational tempo. The target sentences of the present experiment were intermingled with filler sentences from other experiments.

2.3. Analysis

Of the total of 2340 sentences uttered by the 18 speakers (130 each) 63 were not considered in the final results because of non-measurable accents, mostly due to creaky voice. Altogether 2277 sentences were retained for analysis. The recorded sentences were digitized at a sampling rate of 16 kHz. They were analyzed using the acoustic speech analysis software Praat (Boersma & Weenink, 2006). The recordings were partly automatically and partly manually divided into labeled substrings with the help of spectrograms and acoustic inputs. Obvious errors due to the f0 algorithm (for instance octave jumps) were corrected by hand, and the contour was smoothed using the Praat smoothing algorithm (frequency band 10 Hz) to minimize microprosodic perturbations. All frequency measurements were semi-automatically done using a script that detects the highest f0 value within a given domain. The domains for measurements were the complementizer, each argument (article plus noun), the participle and the auxiliary has. An example of the segmentation appears in (5) with ‘#’ as the indicator of a boundary.

(5) # Weil # der Reiher # den Hummer # eingeladen # hat #
# because # the.NOM heron # the.ACC lobster # invited # has #

Table 1
Parameters varied in the production experiment; focused constituents are underlined; altogether 26 conditions

<table>
<thead>
<tr>
<th>All-new</th>
<th>Narrow focus on one argument</th>
<th>Narrow focus on the verb</th>
<th>Focus on all arguments</th>
</tr>
</thead>
<tbody>
<tr>
<td>NV</td>
<td>NV</td>
<td>NAV</td>
<td>NAV</td>
</tr>
<tr>
<td>NAV</td>
<td>NAyV</td>
<td>NAYV</td>
<td></td>
</tr>
<tr>
<td>NDV</td>
<td>NDV</td>
<td>NDV</td>
<td></td>
</tr>
<tr>
<td>NDAV</td>
<td>NDAV</td>
<td>NDAV</td>
<td></td>
</tr>
<tr>
<td>NDV</td>
<td>NDAV</td>
<td>NDAV</td>
<td></td>
</tr>
<tr>
<td>NDAV</td>
<td>NDAV</td>
<td>NDAV</td>
<td></td>
</tr>
<tr>
<td>NDV</td>
<td>NDAV</td>
<td>NDAV</td>
<td></td>
</tr>
<tr>
<td>NDAV</td>
<td>NDAV</td>
<td>NDAV</td>
<td></td>
</tr>
<tr>
<td>NDV</td>
<td>NDAV</td>
<td>NDAV</td>
<td></td>
</tr>
<tr>
<td>NDAV</td>
<td>NDAV</td>
<td>NDAV</td>
<td></td>
</tr>
<tr>
<td>NDV</td>
<td>NDAV</td>
<td>NDAV</td>
<td></td>
</tr>
<tr>
<td>NDAV</td>
<td>NDAV</td>
<td>NDAV</td>
<td></td>
</tr>
<tr>
<td>NDV</td>
<td>NDAV</td>
<td>NDAV</td>
<td></td>
</tr>
<tr>
<td>NDAV</td>
<td>NDAV</td>
<td>NDAV</td>
<td></td>
</tr>
<tr>
<td>NDV</td>
<td>NDAV</td>
<td>NDAV</td>
<td></td>
</tr>
<tr>
<td>NDAV</td>
<td>NDAV</td>
<td>NDAV</td>
<td></td>
</tr>
<tr>
<td>NDV</td>
<td>NDAV</td>
<td>NDAV</td>
<td></td>
</tr>
<tr>
<td>NDAV</td>
<td>NDAV</td>
<td>NDAV</td>
<td></td>
</tr>
<tr>
<td>NDV</td>
<td>NDAV</td>
<td>NDAV</td>
<td></td>
</tr>
<tr>
<td>NDAV</td>
<td>NDAV</td>
<td>NDAV</td>
<td></td>
</tr>
<tr>
<td>NDV</td>
<td>NDAV</td>
<td>NDAV</td>
<td></td>
</tr>
<tr>
<td>NDAV</td>
<td>NDAV</td>
<td>NDAV</td>
<td></td>
</tr>
<tr>
<td>NDV</td>
<td>NDAV</td>
<td>NDAV</td>
<td></td>
</tr>
<tr>
<td>NDAV</td>
<td>NDAV</td>
<td>NDAV</td>
<td></td>
</tr>
<tr>
<td>NDV</td>
<td>NDAV</td>
<td>NDAV</td>
<td></td>
</tr>
<tr>
<td>NDAV</td>
<td>NDAV</td>
<td>NDAV</td>
<td></td>
</tr>
<tr>
<td>NDV</td>
<td>NDAV</td>
<td>NDAV</td>
<td></td>
</tr>
<tr>
<td>NDAV</td>
<td>NDAV</td>
<td>NDAV</td>
<td></td>
</tr>
<tr>
<td>NDV</td>
<td>NDAV</td>
<td>NDAV</td>
<td></td>
</tr>
<tr>
<td>NDAV</td>
<td>NDAV</td>
<td>NDAV</td>
<td></td>
</tr>
<tr>
<td>NDV</td>
<td>NDAV</td>
<td>NDAV</td>
<td></td>
</tr>
<tr>
<td>NDAV</td>
<td>NDAV</td>
<td>NDAV</td>
<td></td>
</tr>
<tr>
<td>NDV</td>
<td>NDAV</td>
<td>NDAV</td>
<td></td>
</tr>
<tr>
<td>NDAV</td>
<td>NDAV</td>
<td>NDAV</td>
<td></td>
</tr>
<tr>
<td>NDV</td>
<td>NDAV</td>
<td>NDAV</td>
<td></td>
</tr>
<tr>
<td>NDAV</td>
<td>NDAV</td>
<td>NDAV</td>
<td></td>
</tr>
<tr>
<td>NDV</td>
<td>NDAV</td>
<td>NDAV</td>
<td></td>
</tr>
<tr>
<td>NDAV</td>
<td>NDAV</td>
<td>NDAV</td>
<td></td>
</tr>
<tr>
<td>NDV</td>
<td>NDAV</td>
<td>NDAV</td>
<td></td>
</tr>
<tr>
<td>NDAV</td>
<td>NDAV</td>
<td>NDAV</td>
<td></td>
</tr>
</tbody>
</table>
The analysis was done in a number of steps. Firstly, a Praat script located f0 maxima in each of the domains, thus in (5), five f0 maxima were identified. Secondly, the result of the Praat script was hand-edited to correct spurious labeling. The authors inspected the tone labels against the f0-track, the substring divisions, an auditory impression, and the spectrogram. Where the more flexible Praat script had assigned an H label that was not in the position where the more narrow criteria above would place it (because of obvious errors due to the algorithm), the label was manually moved. Thirdly, another Praat script recovered the f0-values at the positions of the tone labels as well as the tone labels themselves, and collected them in a table. Finally, for the duration measurements, a Praat script collected the duration of each domain, as shown in (5).

For statistical analyses, the obtained Hz values were aggregated within each participant and each condition. These aggregated scores were subjected to a repeated measures ANOVA in the case of focus and givenness, and to paired-sample t-tests in case of tonal effects, with speakers as random factor. The figures and tables show the aggregated scores per condition averaged across speakers.

3. Results

3.1. All-new sentences

According to our Hypothesis 1 we expected that in the case of wide focus, all arguments would be accented. Depending on the extent of integration with the preceding argument, we further expected the verb to be sometimes accented and sometimes not. Moreover, we expected a downstepped pattern in all four conditions. The first two expectations were fulfilled while the last one has to be revised.

Altogether 348 realizations of all-new sentences were analyzed in the four all-new sentence conditions. All arguments were associated with a pitch accent in all instances of an all-new sentence. Two classes of sentences could be distinguished: an accented verb was realized in 71% (247) of the sentences and an unaccented verb in 29% (101).3 The ratio between the groups accented and unaccented verb across sentence conditions is illustrated in Table 2. As can be seen from these data, the verb was accented more often in a short sentence than in a longer one. All speakers except one (speaker 6) showed free variation in the accenting on the verb (see Appendix A for an overview of the speaker variation).

In (7) and Fig. 2, both versions of a wide focused sentence are illustrated with the same speaker. In (7a), the verb is accented, whereas it is unaccented in (7b). Recall that a sentence that is entirely focused is meant to be uttered as the answer to a question in which neither the participants nor the verb has been mentioned in the preceding context. It differs from a sentence with a narrowly focused constituent in lacking a given (previously mentioned) constituent. From the representative pitch tracks in Fig. 2, it is clearly visible that the verb has a pitch accent in the first case, but not in the second.

(7) Why were the animals happy?
   a. (Weil der Hümmer dem Löwen den Rammler vorgestellt hat)F
      ‘Because the lobster introduced the buck to the lion.’

---

3The large number of accented verbs in the all-new sentences may be an artifact of our experimental design: the focused constituents were underlined. It may be the case that the informants felt compelled to actually accent all underlined words. We suspect that in a natural situation more verbs would be left unaccented.
Turning now to accent scaling, the expectation that the pattern would show downstep everywhere was not met.

Two large classes of cases must be distinguished: In 45.7% of the sentences, a regular downstep took place, in which each accent was lower than the preceding one. But in the remaining 54.3%, raising of an accent was observed. The raised constituent was either on the verb, in which case, of course, the verb was accented (Fig. 2a), or on the preverbal argument, in which case the verb was unaccented (see Fig. 2b). In a pattern involving raising, the last accent interrupted downstep and was realized much higher than predicted by a regularly descending pattern.

Not a single speaker used the downstep pattern in every all-new sentence, but all speakers used it at least once. The same is true for the upstep pattern: Every speaker produced at least one utterance with an upstep on a constituent, but nobody used the upstep pattern consistently (see Appendix A).

Fig. 3 shows the average f0 value for each accent in all patterns across all speakers. First, when the verb was accented, three patterns were identified: downstep throughout, upstep on the preverbal argument, upstep on the verb. Second, when the verb was unaccented, two patterns were realized: downstep throughout, and upstep on the preverbal argument. The values in the cells and on the graphs, if nothing indicates the contrary, are averaged high tones for all speakers and all utterances.

Some values were constant. The unaccented verbs had an average f0 varying between 186 and 190 Hz, and the pitch of the raised argument preceding the unaccented verb was also constant, between 267 and 280 Hz.

---

4In a few cases (15 realizations altogether), the raising took place on the final argument preceding an accented verb.

5There were a few sentences in the corpus where the arguments were realized either at the same height or where the later arguments were pronounced even higher than the early ones. Their number is insignificant, and we did not place them in a separate group. Instead we decided to treat them as lacking downstep, and they appear in the group of sentences with raised arguments.
3.2. Effect of narrow focus on pitch accents

A narrow focus systematically induced a bitonal pitch accent on the narrowly focused argument or verb. Hypothesis 2, which posited that a pitch accent is raised in a narrow focus configuration relative to its value in an all-new sentence, was entirely confirmed.

Fig. 3. Mean f0 of all-new sentences in different realizations: (a–c) accented verb, (d–e) unaccented verb; the percent of occurrences relates to the total number of 348 all-new sentence realizations. (a) regular downstep, accented verb; 33.9%. (b) upstepped argument, accented verb; 4.3%. (c) upstepped verb, accented verb; 32.8%. (d) regular downstep, unaccented verb; 11.8%. (e) upstepped argument, unaccented verb; 17.2%.

3.2. Effect of narrow focus on pitch accents

A narrow focus systematically induced a bitonal pitch accent on the narrowly focused argument or verb. Hypothesis 2, which posited that a pitch accent is raised in a narrow focus configuration relative to its value in an all-new sentence, was entirely confirmed.
Narrowly focused sentences are illustrated in (8) and in Fig. 4. Fig. 4a shows a narrow focus on the nominative, in which case the first argument is accented and all following arguments and the verb are unaccented (8a). Fig. 4b illustrates a narrow focus on the final verb (8b). In this case, all preceding arguments have a pitch accent. In other words, the postnuclear deaccented material in Fig. 4a is much flatter than the unfocused prenuclear material in Fig. 4b.

(8)  
a. Weil (der Löwe) dem Reiher den Hammel vorgestellt hat.  
′Because the lion introduced the sheep to the heron.′

b. Weil der Reiher dem Hummer den Löwen (vorgestellt) hat.  
′Because the heron introduced the lobster to the lion.′

Fig. 5 shows the mean pitch for the sentences with a unique narrow focus (a: A1, b: A2, c: A3, d: V). A pitch accent on a narrow focus was higher than the corresponding accent in an all-new sentence with downstep, as can be gathered from a comparison with the values for A1, A2, and A3 in all-new conditions (the cases that show the unmarked downstep pattern) from Table 3. A three-way repeated measures ANOVA with the factors focus, position of the arguments, and number of arguments confirmed the difference between all-new and narrow focus sentences with a main effect of focus ($F(1,17) = 15.85; p < .001$). A significant interaction of focus and position ($F(1,17) = 11.50; p < .005$ for position one and two; $F(1,17) = 12.93; p < .005$ for position two and three) showed that this difference depends on the position of the argument, the effect being stronger for A2 than for A1 and A3. The interaction of number of arguments and focus showed no significant effect for one- and two-argument structures. Yet a significant difference was found between two- and three-argument structures ($F(1,17) = 9.51; p < .01$), meaning that the scaling of longer sentences differed significantly among focus structures. Finally, the interaction of focus, number of arguments, and position taken together showed a significant effect ($F(1,17) = 10.34; p < .005$), which indicated that the two factors position of the arguments and number of arguments had an impact on the scaling of the accents in different focus structures.

---

*This is not true of the value of the nominative in a NV sentence. The reason is that only the all-new sentences realized with a downstep pattern entered this table. If the results of the sentences with upstep on the verb displayed in Fig. 3c were taken into account as well, the pattern would be the same as in the other conditions.*
The remaining patterns are those in which more than one argument is focused, but the verb is not; see the values of the high tones in Table 4. In conditions NDV and NAV, two arguments are focused (A1 and A2), and in condition NDAV, all three arguments are focused. In the latter configuration, downstep applied between the first two arguments. The difference in height between A1 (275 Hz) and A2 (243 Hz) was on average and across all speakers 32 Hz, and between A2 and A3, 20 Hz. Upstep applied on A3. In the sentences with two arguments, the upstep occurred on the second argument, which was immediately preverbal, and the effect of downstep was obscured or cancelled out.

An interesting observation, shown in Table 3 above, is that an initial narrowly focused A1 was higher in pitch than a narrowly focused A2, which in turn was higher than a narrowly focused A3. More generally, constituents decreased in pitch height from the beginning to the end of the sentence across conditions. This can
be interpreted as an effect similar to downstep. The narrowly focused constituents later in the sentences were preceded by prefocally accented constituents.

Finally, in line with all other constituents boosted by a narrow focus, a focused verb had a higher f0 when it was narrowly focused than when it was part of an all-new sentence or when it was unaccented (see Fig. 5d). However, f0 on an accented verb was definitely lower than f0 on a narrowly focused argument. The narrowly focused verb was only slightly higher or even lower than A1, except for in the NV pattern, in which the verb was much higher than the unique argument.

To sum up this section, a narrow focus had the effect of raising an accent’s f0 locally. A narrowly focused word’s f0 was always significantly higher than the same word in a downstepped pattern of an all-new sentence.

### 3.3. Effect of givenness on pitch accents

#### 3.3.1. Prefocal givenness

Hypothesis 3 was also confirmed. Prefocal givenness lowered f0 of pitch accents, but did not delete them. Perceptually, and also compared with postfocal deaccented constituents, prefocal arguments had pitch accents, although these were lower than their equivalents in a focused context, as is shown in Fig. 6. The f0 of an initial given nominative (the bars C and D in Fig. 6) was lower than an initial nominative in an all-new sentence (A) and in a narrow focus (B). These results show a difference between the two prefocal givenness conditions: f0 of the given nominative was lower prior to a narrow focus than prior to another given constituent. This difference cannot be explained by means of information structure alone, and we come back to these data in Section 3.7 below.

Also sentence-medially, prefocal lowering applies. When A2 was given, its f0 varied on average across all speakers between 224 and 230 Hz. When it was part of an all-new pattern, its f0 was on average between 253 and 257 Hz. And when A2 was narrowly focused, the f0 range was between 271 and 276 Hz.

Table 4
Pitch peak values in Hz on the arguments and the verb in cases where all arguments are focused, but the verb is not; focused constituents are underlined

<table>
<thead>
<tr>
<th></th>
<th>A1</th>
<th>A2</th>
<th>A3</th>
<th>V</th>
</tr>
</thead>
<tbody>
<tr>
<td>NDV</td>
<td>264</td>
<td>270</td>
<td>188</td>
<td></td>
</tr>
<tr>
<td>NAV</td>
<td>257</td>
<td>269</td>
<td>189</td>
<td></td>
</tr>
<tr>
<td>NDAV</td>
<td>275</td>
<td>243</td>
<td>263</td>
<td>183</td>
</tr>
</tbody>
</table>

Fig. 6. Mean f0 showing a 95% confidence interval for the initial nominative in different environments: (A) Nominative in an all-new pattern. (B) Narrowly focused nominative. (C) Given nominative before a narrow focus. (D) Given nominative before a given argument.
As illustrated in Table 5, a sequence of two prefocal given arguments (A1 and A2) showed downstep. The downward steps were about 12 Hz on average when the two given arguments preceded a narrowly focused verb. When the narrowly focused constituent was an argument, the downward steps were larger: around 27 Hz.

### 3.3.2. Postfocal givenness

In the postfocal position, arguments were deaccented, in confirmation of Hypothesis 3. First, Fig. 7 compares the value of the verb in different environments. Columns A and B show f0 of the verb in an all-new sentence and in a narrow focus environment, respectively. Columns C to E show the value of a verb in a given context when the arguments have different focus patterns (see legend). It can be observed from these data that, in the postnuclear position (thus C–E), the verb has a very stable value.

The same deaccenting effect could be seen on postnuclear medial values, which were lower when the constituents were given than when they were new. This is shown in Table 6 for some relevant conditions. The initial constituent is always focused, and the medial argument is given in the first/second columns and focused in the third/fourth ones.

Finally, there is evidence that given constituents are in a downstep relationship to each other. When a second (and third) postnuclear constituent was present, downward steps were still realized, even though they became much smaller at the end of the sentence, as Fig. 5a shows and as Fig. 5b suggests. We do not try to

---

**Table 5**

<table>
<thead>
<tr>
<th>Condition</th>
<th>A1</th>
<th>A2</th>
<th>A3</th>
</tr>
</thead>
<tbody>
<tr>
<td>NAV</td>
<td>249</td>
<td>239</td>
<td></td>
</tr>
<tr>
<td>NDV</td>
<td>251</td>
<td>238</td>
<td></td>
</tr>
<tr>
<td>NDÅV</td>
<td>257</td>
<td>229</td>
<td></td>
</tr>
<tr>
<td>NA&amp;DV</td>
<td>255</td>
<td>230</td>
<td></td>
</tr>
<tr>
<td>ADNV</td>
<td>255</td>
<td>227</td>
<td></td>
</tr>
<tr>
<td>ANDV</td>
<td>251</td>
<td>224</td>
<td></td>
</tr>
<tr>
<td>DANV</td>
<td>256</td>
<td>227</td>
<td></td>
</tr>
<tr>
<td>DNÅV</td>
<td>253</td>
<td>226</td>
<td></td>
</tr>
<tr>
<td>NDÅV</td>
<td>258</td>
<td>233</td>
<td>229</td>
</tr>
</tbody>
</table>

---

Fig. 7. Mean f0 showing a 95% confidence interval for the effect of givenness on the verb. (A) All-new condition. (B) Narrow focus on the verb. (C) Narrow focus on preverbal argument. (D) Narrow focus on pre-preverbal argument (preverbal argument and verb are given). (E) All preverbal arguments are focused.
3.4. Effect of narrow focus on duration

Hypothesis 4 predicts that narrow focus has an effect on duration: a constituent which is narrowly focused is expected to be longer than when it is part of an all-new sentence (see Cooper et al. (1985) for such an effect in English). The hypothesis was confirmed in our data. Narrowly focused constituents were significantly ($F(1,17) = 20.775; p < .001$) longer than non-focused ones, a result shown in Fig. 8. A second result is also apparent from this figure: nominative arguments were shorter than their accusative and dative correspondents. This difference may reflect a difference in their position in the sentence since a nominative occurred initially more often than the other constituents, though more data are needed to confirm this hypothesis. The longer duration of the verb comes about because it is quadrisyllabic (except vor.ge.stellt ‘introduced’), as opposed to the arguments, which are always trisyllabic (article + noun).

A further effect is that duration can be related to the number of constituents that a sentence contains. Fig. 9 shows that the duration of an initial nominative increased significantly ($F(2,34) = 21.60; p < .001$) when followed by one, two and three constituents (verb included). The increase in duration between one and two arguments was larger than the one between two and three arguments. We have no explanation for the fact that the realization of the nominative takes more time when there was a greater number of following constituents. As a reviewer suggests, the greater number of prosodic phrases might exert an influence here, but in a way that we do not understand.

A reviewer observes that this effect looks like declination in the sense of Pierrehumbert (1980), i.e. a phonetic lowering effect independent of phonological downstep. An alternative explanation could be that downstep takes place in a compressed pitch range.
3.5. Word order

The last hypothesis related to information structure was on word order: a non-canonical word order (in which the nominative is not initial) may have an influence on the height of pitch accents (Hypothesis 5). This hypothesis was not confirmed: word order did not affect any of the values calculated in this study. As far as pitch is concerned, we compared the accents of the arguments in the unmarked patterns NAV and NDV with those in the marked patterns ANV and DNV. Sentences with the same arguments but in a different order did not differ in the peak f0 of their arguments. The f0 for A1 and A2 were not significantly different regardless of their nominative, accusative or dative case and regardless of their information status. In the longer conditions, again no difference could be found in the f0 of accented words. The values were similar in unmarked and in marked word orders.

In sum, no significant effect of word order was found, neither in the focused constituents nor in the given ones. As this could be an artifact of the experiment, in which informants spoke in a rather mechanical way, we consider this issue as unresolved.

3.6. Constant final value and final f0 drop

Hypothesis 6 predicted a constant final value for all our data, regardless of length and information structure. This hypothesis was confirmed by the data, and is thus in line with Liberman and Pierrehumbert (1984) and Grabe (1998). Figs. 5a–c and 3d and e illustrate the constant final low point of the speakers in our data. Only the values on an unaccented verb can be used here. Only in these cases, the maximum f0 measured by the Praat script corresponds to the lowest value of the sentence.

A much larger downward trend than predicted by downstep between the last narrowly accented argument and the first unaccented constituent was observed. This considerable drop in f0 was on average 80 Hz. Interestingly, not only the final low of the sentence but the drop itself was constant: 80–85 Hz. Paired-sample T-tests on the comparison of the final drop across different conditions showed no significant difference between narrow focus in different positions. But the low f0 value that was reached depended on which of the first, second or third arguments was narrowly focused. At the end of the sentence, the constant final value of 190 Hz or less was reached in all cases, but sometimes it needed more than one decreasing step to be achieved (see the values in Fig. 5a and b).

We also find a constant final value in all-new sentences in the realizations with an unaccented verb, although there was a significant difference between the all-new condition and the narrow focus conditions. The final drop was significantly larger for narrow focus ($t(17) = 6.58; p < .001$ for all-new vs. initial narrow focus; $t(17) = 7.79; p < .001$ for all-new vs. second argument focus; $t(17) = 7.60; p < .001$ for all-new vs. third argument focus). The difference is due to the fact that the all-new sentences had different realizations, only some of which involved final f0 drop.\(^8\)

\(^8\)We only found the final drop for sentences containing an unaccented verb, thus 101 instances out of a total of 348 all-new sentences.
3.7. Further tonal effects

Following Hypothesis 7, all pitch values of high tones in our data could be interpreted as a combined effect of syntactic structure and focus/givenness partition. Our data cannot confirm this hypothesis. Some values that were calculated on the tones themselves, as for example the raised values on the upstepped argument in the all-new patterns, or the different values of a given argument which vary as a function of the adjacency or non-adjacency of the narrowly focused constituent, could not be explained by syntax and information structure alone. In other words, the f0 on some pitch-accented words could be explained neither in terms of regular downstep, nor by boosting due to narrow focus, nor by lowering due to givenness.

The first such value concerns a narrowly focused pitch accent just before the final drop. In this environment, the pitch was raised, an effect which is not limited to the preverbal accent, but also affects an accent located immediately before an unaccented constituent, regardless of both sentence position and information structure.

Table 7 makes some relevant comparisons from which it appears that the f0 of an accented argument is lower when it precedes a further accented argument (second column) than when it precedes an unaccented one (fourth column). We interpret the higher pitch peaks on the latter case as a result of H-tone raising. On average, the difference between a raised tone and a non-raised one was 17 Hz and was significant (t(17) = 4.12; p < .01 for the first argument; t(17) = 5.03; p < .001 for the second argument).

N of NDAV, i.e. a nominative followed by two more arguments, did not show a significantly raised f0 when the following dative was unaccented. In this case, there was only 5 Hz difference between the value of N in an all-new context and in a narrow focus. The second place where the difference was only weak is A3 in NDAV. Some of the raised values in Table 7 are indistinguishable from the effect of narrow focus. But two facts cast doubt on the view that raising of a high tone can always be explained as a result of narrow focus. First, as is visible from Table 7, A2 is raised both when it is narrowly focused (e.g. NAV), and when it is part of a pattern in which all arguments are focused with the exception of the verb (e.g. NAV). Second, H-raising also takes place in an all-new pattern, when the verb is unaccented and the preverbal argument is upstepped. The relevant values from Fig. 3a and d are reproduced in Table 8. The upstepped argument is on average about 20 Hz higher than in a downstep pattern. In this case, narrow focus plays no role at all.

The explanation advanced here is that of a purely tonal interaction between two adjacent tones. We think that the phenomenon of H-raising, which typically occurs when a high tone precedes a low tone in tone

---

We suspect that a ceiling effect is at play in this case, but more investigations are needed to confirm this tentative explanation.

---

<table>
<thead>
<tr>
<th>Table 7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pitch peak values on the first (A1), second (A2), and third (A3) arguments before focused constituents in all-new sentences (left column) and before non-focused, i.e. given, constituents (right column); in narrow focused sentences, underlined constituents are focused, the others given.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Before [+foc]</th>
<th>Before [-foc]</th>
</tr>
</thead>
<tbody>
<tr>
<td>A1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NV</td>
<td>264</td>
<td>NV</td>
</tr>
<tr>
<td>NDV</td>
<td>271</td>
<td>NDV</td>
</tr>
<tr>
<td>NAV</td>
<td>269</td>
<td>NAV</td>
</tr>
<tr>
<td>NDAV</td>
<td>284</td>
<td>NDAV</td>
</tr>
<tr>
<td>A2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NDV</td>
<td>257</td>
<td>NDV</td>
</tr>
<tr>
<td>257</td>
<td>NDV</td>
<td>275</td>
</tr>
<tr>
<td>NAV</td>
<td>255</td>
<td>NAV</td>
</tr>
<tr>
<td>255</td>
<td>NAV</td>
<td>271</td>
</tr>
<tr>
<td>NDAV</td>
<td>253</td>
<td>NDAV</td>
</tr>
<tr>
<td>A3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NDAV</td>
<td>260</td>
<td>NDAV</td>
</tr>
<tr>
<td>260</td>
<td>NDAV</td>
<td>269</td>
</tr>
</tbody>
</table>
languages (see among others Connell & Ladd, 1990; Laniran & Clements, 2003; Xu, 1997 for H-raising in tone languages) plays a role in German as well.

The second values, which cannot be explained by syntax and information structure, can be understood as the reverse effect of H-raising. A high tone immediately preceding a raised high tone was lowered as compared with the same tone preceding a non-raised high tone. Several data sets illustrate this effect.

First, as was shown in Fig. 6 in Section 3.3.1, the lowest values for an initial (given) nominative were those in which it immediately preceded a narrow focus (column C). When the second argument was given as well (column D), the values of the nominative were higher by ca. 20 Hz on average across speakers. Paired-sample T-tests revealed that a comparison between the values of the two initial given nominatives produced a significant effect ($t(17) = 8.50; p < .001$). This difference cannot be explained by any other effect than a purely tonal one. In particular, it cannot be explained if only information structural effects are taken into account, since the given status is identical in both cases.

Second, Table 9 compares the height of arguments immediately preceding a narrowly focused argument with similar ones in the downstep pattern of all-new sentences. The former values (left column) were surprisingly stable (on average between 229 and 239 Hz), and systematically lower than the values they had in an all-new sentence (right column).

Even if it is assumed that givenness is responsible for the low values in the left column of Table 9 such an interpretation is not available for the third set of data: an argument preceding an H-raised tone in an all-new sentence. The relevant values are shown in Table 10. Paired-sample T-tests show that, in an all-new sentence, a constituent preceding an accented constituent was significantly lower than a corresponding constituent in a regular downstep pattern ($t(17) = 7.06; p < .001$). The difference was again 17 Hz or more on average. In this case, as well, an argument preceding a raised value was lowered as compared with a downstep pattern. Since the all-new status of the constituents was identical in each case, it must be assumed that it was the tonal makeup of the phrase that was responsible for the difference.\(^{10}\)

4. Discussion

The data on tone scaling presented in Section 3 confirm a view of the tonal structure of German as largely determined by syntax and information structure. But they also revealed that syntax and information structure

---

\(^{10}\)The discussion of Table 7 is also relevant. It was shown there that downstep between A1 and A2 before a raised argument is larger than downstep before a narrowly focused verb. In the latter configuration, no lowering can take place, because the verb is never H-raised (see discussion).

---
are not sufficient to explain all patterns of variation, and we proposed to add a third component to pitch accent scaling: interactions between the tones themselves. In this section, it is shown how the three components entering tonal scaling are integrated into a model of German intonation.

The prosodic phrasing of the experimental sentences is illustrated in (9) and (10). Every argument projects its own p-phrase. An all-new sentence can integrate the verb in the preceding p-phrase, as in (9a), or the verb forms a separate p-phrase, as in (9b) (see Fuchs, 1976; Gussenhoven, 1992; Jacobs, 1993). In such a sentence, every p-phrase has a head, realized as a pitch accent.

(9) a. [Weil der HAMMEL]P [den HUMMER eingeladen hat]P

because the sheep the lobster invited has

‘Because the sheep invited the lobster.’


In (10), a sentence with narrow focus and postfocal deaccenting is shown. Subscript F stands for focus, and G for givenness is present. Two phrasing structures are possible. Phrasing (10a), consisting of only one prosodic phrase, respects the assumption that every prosodic phrase should have its own head. Since the object and the verb are deaccented, they do not form a prosodic phrase, and are integrated into the phrase of the accentuated subject. In (10b), the phrasing based on syntax is kept. Only the height of pitch accents is changed, which are compressed in the second phrase. We have no clear argument that would allow us to choose between the two structures. Further research is needed on the issue of the phrasing of postfocal material, and on the nature of the deaccenting.

(10) a. [(Weil der HAMMEL)F (den Rammler eingeladen hat)G]P


‘Because the sheep invited the buck.’

The tonal pattern of the sentences as illustrated in (11) contains prenuclear rising accents, which are transcribed as bitonal L* Hp sequences, and nuclear falling accents which are transcribed as H* Lt. L* is a low tone associated with an accented syllable, and H* is the high pendant. Hp is the boundary tone of a prosodic phrase (p-phrase), and L1 the boundary tone of an intonation phrase (i-phrase). Autosegmental–metrical transcription was originally proposed by Pierrehumbert (1980) for English (but see Féry, 1993; Grabe, 1998; Grice, Baumann, & Benzmüller, 2005; Truckenbrodt, 2002 among others, for adaptation to German). A low boundary tone (Lt) is associated both with the syllable following H* and with the end of the sentence. The low stretch characterized by this tone extends backwards from the end of the sentence to the nuclear accent. This is in line with the OT implementation by Gussenhoven (2004) for English.

\[
\begin{array}{cccc}
L^* & H_p & L^* & H_p & H^* & L_1 \\
\end{array}
\]


‘Because the buck introduced the heron.’

\[11\] Note that the transcription conventions differ considerably between researchers due to different systems of intonational phonology. See Kügler (2007, 25ff.) for a thorough discussion of different approaches to German intonation.
We consider regular downstep as the unmarked tonal realization (e.g. Féry, 1993; Truckenbrodt, 2002, 2004), in which unmarked prosody causes relatively small but regular descending steps, on average between 17 and 29 Hz. This pattern appeared in approximately half of the all-new sentences of our corpus ($n = 159, 45.7\%$), as discussed in Section 2.1. Downstep also appeared in further environments: in a sequence of two prefocal pitch accents (see Section 3.3.1), and between A1 and A2 when all the arguments are focused with the exception of the verb (Section 3.2). In short, downstep features regularly in a sequence of two or more arguments when they have the same informational status, i.e. when they are equally focused or when they are equally given. But downstep can be smaller or larger depending on the tonal context in which it occurs. For given constituents, the downstep steps are smaller since the scaling of the pitch accents is already reduced due to givenness, while for focused constituents these steps can be larger due to the expanded pitch range.

We analyze the high part of a rising accent $L^*H_P$ as the boundary tone of the p-phrase it ends; see (11) above for an illustration. The high parts of the rising contour are maximally as high as the reference line (top line) of their own domain. Downstep creates new (and downstepped) reference lines reached by the high tones. We assume that downstep takes place across p-phrases, as well as inside of p-phrases, in disagreement with Beckman and Pierrehumbert (1986), but in line with Bruce (1977), Clements (1990), Ladd (1990), van den Berg, Gussenhoven, and Rietveld (1992), Truckenbrodt (2002), Féry and Truckenbrodt (2005), among others (see Féry & Ishihara, to appear, for an elaborate model of German prosody along these lines). The register top and bottom lines are the intervals of the speaker’s voice range, within the limits of which pitch excursions are scaled. Thus, when a high boundary tone is raised, it means that it is the register of the p-phrase delimited by this boundary tone that is raised. An alternative explanation for downstep has been advanced by Pierrehumbert (1980) and Liberman and Pierrehumbert (1984) for intonation languages, in which a medial L tone affects a following H in such a way that the affected H is lower than a preceding H. This effect takes the form of a left-to-right progressive assimilation. It is assimilatory because a low tone lowers a following high tone, thus rendering it more similar to itself. However, the assimilation-driven account of downstep cannot explain instances of downstep with no intervening low tones, as is the case in our data, as in (11) for example, where the high phrasal boundary tone $H_P$ and the high tone of the pitch accent $H^*$ are in a downstep relationship to each other. Assuming downstep as a register effect related to prosodic domains includes cases such as (11). Further, downstep as a register effect taking place in different prosodic domains allows to account for embedded effects of downstep, as well as upstep.

The effect of information structure on the height of accents and boundary tones was discussed in Sections 3.2 and 3.3. The value of the high boundary tone of non-final p-phrases, as well as the value of the high pitch accent of the final p-phrase were measured and compared to each other. It was found that information structure affects the top lines of prosodic domains. Narrow focus raises it, and as a consequence the high part of $L^*H_P$ is higher, and givenness lowers it prefocally and compresses it considerably postnuclearly. The high tone values were systematically higher in a constituent with a narrow focus than in its correspondent with an all-new context or in a given context. The effect of downstep was never cancelled, as a comparison between the values of a narrowly focused first, second and third arguments revealed. In short, the combined effects of syntax and information structure go a long way in the explanation of the high tone values of our data. However, other effects were acting on the scaling of tones, as well, to which we turn in the remainder of the discussion.

The final accent of the declarative sentences examined in the experiment was always a falling tone, which we represented with a high starred tone followed by the low boundary of the i-phrase, $H^*L_I$. The boundary tone $L_I$ reached the bottom of the speaker’s voice, a value we have called the ‘constant final value.’ In a sentence containing postfocal material, the final level was reached immediately after the last accent, or nearly so, otherwise it was attained on the final word of the sentence.

Final drop takes place in a considerable number of cases. This effect characterizes the steep fall between the last and low boundary tone $L_I$ and the preceding upstepped high part of the pitch accent $H^*$. It was regularly found in narrow focus, but also in 54% of the all-new sentences. We assume that the low tone $L_I$ has a dissipimatory effect on a preceding high tone, the dissimilation being obligatory when the high tone stands for a narrow focus, and optional in an all-new sentence. The right-to-left dissimilation between the $L_I$ and the $H^*$ causes an important difference in pitch. We found that the decrease in Hz taken by this final drop was
remarkably stable. Since final drop happens independently of the information status of the constituents carrying these tones, it cannot be considered as an effect of focus.

Final drop should not be confused with Final Lowering, an effect which has been abundantly described in the literature on intonation languages (see Liberman & Pierrehumbert, 1984; Prieto, Shih, & Nibert, 1996; Truckenbrodt, 2004 for very interesting and elaborate analysis of Final Lowering in German). Final lowering affects the last pitch accent in a series of accents, the final step is larger than expected by exponential decay of the decreasing values of the sequence of high tones. However, final drop is a different phenomenon, which relates to the register reference lines and not to accents. The steeper fall involved in final drop is not the last one in a series of accents, but it is the first in a series of unaccented words, or the only unaccented word, and the drop in f0 is much larger than the last step of a series of accents. The only places in our data where it would make sense to search for Final lowering are the sentences in which all constituents are accentuated, even the verb, since this is the only constellation in which downstep occurs with complete regularity. Yet, the effect of final lowering lies not in the scope of this paper, and thus we do comment on this issue.

Phonologically, we analyze the high value of the last H* before final drop as a dissimilation between a high tone and a following low tone. At the end of the intonation phrase, the final L1 exerts a local raising effect on an immediately preceding H*. The process of H-raising appears to take place in two steps. First, the final low tone, L1, is scaled at the bottom line of the voice register, or close to the bottom, if it is not the last acceptable constituent of the sentence. This L1 is also aligned with the syllable following the nuclear accent, creating in this way a flat and low postnuclear contour. The second effect is the increase in pitch that L1 causes on the preceding H*, which then raises its value independently of the tones preceding it. It was shown in Section 3.7 that the value of the raised H* is dependent on final drop, as it is the difference between H* and L1 which matters.

The analysis of H-raising as a dissimilation process between two tones is supported by studies uncovering a similar effect in a number of tone languages. Xu (1997) examines anticipatory and carry-over effects of Mandarin tones, and finds that a high tone is higher when the following tone is low. He concludes that a dissimilatory effect is at play. Similarly, Laniran and Clements (2003) find a dissimilatory H-raising effect in Yoruba: a low tonal onset raises the f0 of the preceding high tone (see also Connell & Ladd, 1990). Gandour, Potisuk, and Dechongkit (1994) also find an H-raising effect in Thai.

Alternative explanations propose that boosting of an accent is always an effect of narrow focus (Cooper et al., 1985; Eady & Cooper, 1986; Eady et al., 1986; Liberman & Pierrehumbert, 1984 for English; Kubozono, 1989 for Japanese). Clearly, our data refute this thesis since we find the same raising effect in all-new sentences. In the German sentences examined in this paper, H-raising happens on a narrow focus as well as in half of the realizations of an all-new sentence. As far as narrow focus is concerned, H-raising is completely regular, and this in all positions and in all sentences.

In their study of tonal interactions in Yoruba, Laniran and Clements (2003) show that high tones are sometimes higher than predicted. First, H-raising happens before L, and second, register adjustment may take place, like reset of high tones in a long sequence of high tones or raising of an initial H in anticipation of a long high tone sequence. They express their proposal in the following terms: ‘[...] although Yoruba speakers implement downstep and H-raising by quantitatively different means, their realization strategies ‘conspire’ to insure that downstepping H tones will not penetrate the frequency band reserved for M tones’ (2003, p. 232). In other words, H-raising guarantees that a downstepped H is not mistaken for a mid or even a low tone. However, their explanation for H-raising does not fit our data. Our speakers do not reset their voice in the same way as Yoruba speakers do. It is not clear whether there is a tonal domain reserved for low tones in German.

Upstep, as described by Truckenbrodt (2002), is not adequate for our data either. This phenomenon is limited to boundary tones of intonation phrases and it only affects medial i-phrases. The effect described here is a different one: it affects sentence-final bitonal accents, analyzed as pitch accents followed by boundary

---

12 In a nutshell, Truckenbrodt proposes that final lowering is the effect of absence of raising of a high tone before a downstep tone. Since a sequence of high tones are in a downstep relationship, all are affected except for the last one.

13 This may be a simplification, due to the observation that a non-final postnuclear word may be followed by decreasing steps, but this does not bear on the argumentation in this section.
tones of prosodic phrases. A last difference between H-raising and upstep as analyzed by Truckenbrodt (2002)
is the height of the raised H. The high tone in H-raising does not necessarily return to the value of the first
tone, but rather reaches a point in the speaker’s voice at which it can trigger the final drop. In our data, it is
located at roughly 80 Hz above the point reached by the final drop. The raised H is, in other words, not scaled
relative to the reference line of the first accent in the sentence. A more accurate approximation is that it is
scaled to the last boundary tone, i.e. the constant final low. More research is needed to understand this
phenomenon better, especially when final drop happens relatively early in the sentence.

The raised H* also has a dissimilatory effect from right-to-left on a preceding H, decreasing its value. The
affected high tone is the second tone of a rising bitonal contour that is phonologically represented as L*H_p. In
other words, a starred H lowers the value of the preceding boundary tone of a p-phrase. This effect can be
phonologically analyzed as a regressive dissimilation of an H* on a preceding H_p. The first high tone is
rendered less similar to the second one. As in final drop and in H-raising, H-lowering takes place both when
the raised H* is the narrow focus and in an all-new sentence. The sequence L*H_p H* L_i is the final sequence of
tones in a declarative sentence.

This H-lowering can again be considered as a two-step process. First, the second high tone (the one acting
on the preceding tone) is raised. Second, the first H tone (the one on which the action is exerted) is lowered and
is thus dissimilated from its following tone.

Liberman and Pierrehumbert (1984) observed that a high tone preceding a boosted high tone is lowered, but,
in contrast to our analysis, attribute this effect to information structure. We find, however, that the decisive
factor is not so much the information structure, but rather the tonal structure of the sentence. H-lowering
happens before a raised H independently of the status of the lowered tone as given or as part of an all-new
sentence, and independently of the status of the raised tone as narrowly focused or as part of an all-new pattern.

Thus H*-raising increases at the same time the difference between this last H* and the preceding high tone.
But, when the final verb is focused, no such effect occurs, and the difference between the last tone and the
preceding one is not so large. A final narrowly focused verb has a lower pitch accent (ranging on average
between 255 and 263 Hz) than a narrowly focused preverbal argument (268–271 Hz). The reason for the
different behavior of the final verb is due to the fact that the verb, being the final constituent, is not followed
by a deaccented constituent, and that a final drop is not needed in this context. There is no final drop, and
consequently no H-raising. As a result, H-lowering does not apply either.

A final remark on the effect of different sentence lengths on the pitch scaling is in order. The average values
of an initial nominative in a whole-focused sentence, and in a narrow focus are shown graphically in Fig. 10.
They are clearly higher when the sentence contains more arguments.\footnote{This effect can be compared with
long-distance anticipation, which has been described for tone languages (Stewart, 1993 for Dschang
and Ebrïï, Rialland & Somé, 2000 for Dagara, to cite just a few) and to a lesser extent for intonation languages (see Thorsen, 1985 for
Danish, and the rejection of preplanning by Liberman & Pierrehumbert, 1984; recently for Romance languages, Prieto, D’Imperio,
Elordieta, Frota, & Vigário, 2006). In our data, we find evidence for soft preplanning in the sense of Liberman and Pierrehumbert (1984).} In all-new sentences, there is a
significant raising of the value of the initial nominative when the number of arguments increases. When the nominative is narrowly focused, this value rises between one and two arguments, but it does not increase significantly between two and three arguments.

We tentatively explain the absence of raising in the latter case in the following way: speakers have a register top line which cannot be raised ad libitum. In our data, a longer sentence or a raised nominative already reaches the ceiling. As a consequence, it is not possible to raise this top line even more. A confirmation of this view was noticed in Section 3.3. Raised initial nominatives because of narrow focus are significantly higher than the initial nominative in the all-new pattern, with the exception of NDAV, where the difference is not significant. First a raised nominative was only 5 Hz higher than a non-raised one, and second, it was at the same height as a nominative in a shorter sentence. Also in this case, a ceiling effect may be assumed that blocks further tone raising. However, this has to be shown experimentally.

5. Conclusion

Laniran and Clements’ (2003) study of tonal interactions in Yoruba has significantly influenced the interpretation of the results of the present paper. Laniran and Clements report the results of production experiments in Yoruba and give a careful analysis of the speech of four speakers. They propose that the overall shape of an f0 contour is the result of a compromise between different tendencies in the tonal pattern. The authors give a compositional analysis of tone scaling in which the melodic contour of an utterance is the result of a number of interacting factors. In the same way, we propose that the tonal scaling of high tones in a declarative German sentence is the result of different influences coming from syntax, information structure and tonal dissimilations.

The present work has examined tonal issues in German with the help of a production experiment in which 18 speakers uttered a total of 2277 sentences of the same syntactic structure, but with a varying number of constituents, word order and focus-given structure. This experiment has shown that the scaling of high tones, and thus the entire melodic pattern, is influenced by information structure as has been suggested before for other Germanic languages. However, this experiment also showed that not all the data can be explained by information structure. Therefore, we propose an additional explanation to information structure, namely, that tones act on each other resulting in dissimilatory tonal effects. As for the effects of information structure, they are not especially surprising, as German, like English, is an intonation language whose use of accents and accent sequences is conditioned by pragmatic considerations. For the sentences used in the experiment described here, the influence of information structure can be summed up in the following way: focus raises tones while givenness lowers them in prenuclear position and cancels them out postnuclearly. These changes in the values of accents were explained by the influence information structure has on reference lines associated with prosodic domains. However, and crucially, information structure or normal syntactic structure cannot explain all tonal patterns gathered in the analysis of the 2277 pitch tracks. Further effects appear to be purely tonal ones. First, regular downstep is considered as the default pattern of high tone implementation in a ‘neutral’ sentence, where neutral is understood as all-new or whole-focused. It can be understood as a tonal effect taking place across p-phrases: an L anticipatorily assimilates a following H tone, and thus lowers its value, but we have preferred to analyze it as an effect of downstepping the prosodic reference lines relative to each other. Second, final f0 drop is the steep fall from a raised high tone to the bottom line of the speaker and is found in situations of narrow focus, as well as in all-new sentences, when the last high tone is H-raised. Third, H-raising, which has been described among others by Connell and Ladd (1990), and Laniran and Clements (2003) for Yoruba, and Xu (1997) for Mandarin Chinese, plays a crucial role as a component of the tonal pattern of German. Optional H-raising is triggered by a low tone exerting a dissimilatory regressive effect on a preceding high tone. We argued that raising a last accent allows the phenomenon of final f0 drop to be perceptively prominent. Third, the reverse phenomenon, H-lowering before a raised H, is also at play. This effect lowers a high tone before a raised high tone, and has been analyzed as another dissimilatory regressive effect.

In the discussion, alternative accounts, like upstep, a phenomenon which raises tones at the end of a medial phrase (see Truckenbrodt, 2002), and final lowering, the larger final step in a series of downsteps (see Liberman & Pierrehumbert, 1984; Truckenbrodt, 2004 among others), are shown to be different effects. As far
as upstep is concerned, all our phrases were final, since we avoided syntactic complexity, so that the observed prefinal raise cannot be the result of anticipating a following intonation phrase. And clearly, final lowering is not at play in our data, if final lowering is understood as a larger downstep than predicted by the preceding downward steps. In a downstep pattern, the final step was not larger than the preceding ones. The effect of final f0 drop cannot be understood as final lowering as it involves a preceding raise on the last accent.

A side issue of this paper concerns the question of preplanning (or ‘foresight’). It was shown that speakers expand their register when they know that more accents (and more downsteps) are to come. The dissimilatory effects of H-raising and H-lowering can also be considered as providing some evidence for a mild form of preplanning, as a tone has an influence on a preceding one. H-lowering was shown to even involve a kind of double preplanning. In any case, our data appear to support the idea of ‘soft preplanning’ as argued for by Liberman and Pierrehumbert (1984).

Our investigation has necessarily been limited in scope since it investigates only sequences of pitch accents in a single intonation phrase. It is evident that the study of more complex intonation patterns will provide a clearer picture of tonal interactions.

Acknowledgments

This paper is part of the project A1 of the SFB 632 on Information Structure and of the project Prosody in Parsing, both financed at the University of Potsdam by the Deutsche Forschungsgemeinschaft. Thanks to Gisbert Fanselow, Jonathan Harrington, Robin Hörnig, Shin Ishihara, Ineke Mennen, Anja Gollrad, Gerrit Kentner, Lisa Selkirk, Hubert Truckenbrodt, Thomas Weskott, Wang Bei, Sabine Zerbian and two anonymous reviewers for helpful discussions and comments. The usual disclaimers apply. Our deep gratitude goes to Anja Arnhold, Kirsten Brock, Susanne Genzel and Esther Sommerfeld for technical support.

Appendix A

An overview of the speaker variation is shown in Table A1.

Table A1
Distribution of the realization of all-new sentence patterns by speakers; each of the four conditions were repeated five times ($5 \times 4 = 20$). Less than 20 implies that some f0 values could not be measured due to creaky voice.

<table>
<thead>
<tr>
<th>Speaker</th>
<th>Accented verb Downstep</th>
<th>Raising on argument</th>
<th>Raising on verb</th>
<th>Unaccented verb Downstep</th>
<th>Raising on argument</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>15 2</td>
<td>1</td>
<td>1</td>
<td>1 1</td>
<td>20</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>1 1</td>
<td>6</td>
<td>4</td>
<td>1 9</td>
<td>18</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>2 2</td>
<td>8</td>
<td>4</td>
<td>3</td>
<td>19</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>3 1</td>
<td>7</td>
<td>5</td>
<td>16</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>6 9</td>
<td>1</td>
<td>4</td>
<td>20</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>6 3</td>
<td>11</td>
<td>20</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>4 7</td>
<td>3</td>
<td>1</td>
<td>16</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>2 13</td>
<td>1</td>
<td>3</td>
<td>20</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>9 1</td>
<td>2</td>
<td>5</td>
<td>3</td>
<td>20</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>15</td>
<td>3</td>
<td>2</td>
<td>20</td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>8</td>
<td>8</td>
<td>3</td>
<td>1</td>
<td>20</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>11</td>
<td>8</td>
<td>1</td>
<td>20</td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>12 1</td>
<td>3</td>
<td>3</td>
<td>20</td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>14 1</td>
<td>4</td>
<td>1</td>
<td>20</td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>2 1</td>
<td>2</td>
<td>7</td>
<td>6</td>
<td>18</td>
<td></td>
</tr>
<tr>
<td>16</td>
<td></td>
<td></td>
<td>3</td>
<td>3</td>
<td>20</td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>4 2</td>
<td>3</td>
<td>11</td>
<td>20</td>
<td></td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>4 8</td>
<td>1</td>
<td>6</td>
<td>19</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
References


Féry, C., & Ishihara, S. (to appear). How focus and givenness shapes prosody. In M. Zimmermann, C. Féry (Eds.), *Information structure from different perspectives. OUP.*


The role of duration as a phonetic correlate of focus

Frank Kügler

Institute of Linguistics
University of Potsdam, Germany
f.kuegler@googlemail.com

Abstract
The aim of the present paper is to collect quantitative data on durational patterns of target words in different information structure conditions. The acoustic data will be the basis for the manipulation of stimuli to conduct perception tests that will deal with the role of duration as a correlate of focus prominence. For that purpose, a production study is presented that investigates the effects of information structure (wide, narrow, and contrastive focus, prefocal and postfocal givenness), sentence length and position in a sentence on duration of a target constituent in German. Duration of target constituents was measured in 400 utterances produced by 10 speakers. The predictions that focus increases target word duration has been confirmed, while the expected decrease in duration due to givenness has only been confirmed for prefocal given constituents. Postfocally, duration is equivalent to wide focus duration. The effects of sentence length and position have only partly been confirmed; a constituent seems to be shorter in longer sentences than in short ones, and target words occurring early in a sentence appear to be longer than late occurring ones.

1. Introduction
Communication is a process in which a speaker transfers information to a hearer. In this transfer, any utterance has a certain information structure. Any utterance contains at least a focus [7]. According to [7], an utterance may further contain a non-focus part that may branch in a topic and/or a background. Focus is here understood as the presence of alternatives that are relevant for the interpretation of linguistic expressions, and topic as a referent or constituent which the remainder of the sentence is about [16]. Backgrounded information in this experiment refers to given information, defined as previously mentioned in the discourse [2, 3].

Based on the assumptions of Focus-Prominence-Theory (e.g. [17, 18]) the expression of abstract focus prominence is language specific. Many languages use F0 as a prominent cue. In these languages duration almost ever accompanies focus prominence. This paper is the first step in an investigation to disentangle the role of F0 and duration as phonetic correlates of focus prominence. The general question behind is whether intonation languages like German employ the correlate of duration as a consequence of using higher or later F0 peaks, or whether duration is a functionally relevant prosodic cue on its own.

As has been reported in the literature, focus causes an increase in duration on focused constituents, see for instance [1] for Swedish, [8, 9, 10] for English, [4, 11] for German, and [15] for Korean. In addition to the fact that duration increases in focus, the duration of a constituent also appears to depend on its position in a sentence [11, 12]. Longer durations are found in earlier positions of the sentence. Also, if sentences are longer, i.e. contain more words, the duration of constituents appears to be longer in earlier positions [11].

The relation of duration to different focus domains has been investigated by [4] who found a negative correlation with the size of the focus domain, i.e. the smaller the focus domain the longer the duration of a particular constituent. In general, focus seems to be closely connected to longer duration. This can also be seen in situations where no other tonal correlates of focus are present, i.e. in the case of second occurrence focus (SOF) [5, 13, 14]. In English and German constituents that are in the scope of a focus operator such as ‘only’ but do not receive the nuclear stress of the sentence and thus are not associated with pitch accents show longer durations as equivalent given constituents without any focus particle.

This paper reports on durational patterns in German. The aim of this paper is to provide quantifiable data of the effects of ‘position’, ‘length’, and ‘information structure’ (focus and givenness). Based on these data perception experiments are planned that will investigate the role of duration and focus in more detail. For the purpose of the production experiment sentences were created which contain a target word in different positions in the sentence. In addition, sentences varied in total length (counted as number of syllables per sentence). The sentences were embedded in contexts to elicit different information structures (see below).

2. Method

2.1. Speech materials
Two target words were embedded in carrier sentences of different length in different positions. The two target words are nonsense words, contain only sonorant segments in order to ensure easy pitch tracking. One of the words is monosyllabic and one is bisyllabic with word stress on the first syllable. These nonsense words were used as surnames in order to make their occurrence more natural. The target words are Mohn [ˈmo:n] and Liehner [ˈliː.na:r]. The carrier sentences are of the following structure.

(1) a. Frau Mohn will ein Lamm malen.
   b. Ein Lamm will Frau Mohn malen.
   c. Frau Mohn will ein Lamm im Berliner Tierpark malen.
   d. Im Berliner Tierpark will Frau Mohn ein Lamm malen.

    Two target words were embedded in short (1-a), (1-b) and long sentences (1-c), (1-d) either early (1-a), (1-c) or late (1-b),(1-d) in the sentence. The set of sentences in (1) was put into five different contexts yielding five different information structures of the target words.

The information structural baseline is considered to be an all-new sentence, with no particular part of the sentence in fo-
cus, i.e. the answer of a question like (2-a). Narrow focus was elicited by (2-b), and contrastive focus by (2-c), where Drahner is contrasted with the target word Liehner. The variable of givenness was tested in prefocus and postfocal position. The corresponding questions were either asking for a constituent after the target word, prefocus givenness (2-d), or before the target word, postfocal givenness (2-e).

(2) a. What happened?
   b. Who does want to paint a lamb?
   c. Does Mrs Drahner want to paint a lamb?
   d. Does Mrs Mohn want to paint a hoarse?
   e. Does Mr Mohn want to paint a lamb?

A total of 8 unique question-answer pairs (2 items \( \times \) 2 positions \( \times \) 2 sentence lengths) were constructed and each pair was realized in the 5 information structure conditions, resulting in \( 8 \times 5 = 40 \) sentences per speaker. All the 40 sentence-pairs were presented to each speaker in a pseudo-randomized manner; items from four other unrelated experiments were interspersed as fillers. Four pseudo-randomized lists were prepared to minimize order effects.

2.2. Recording procedures

The experiment was carried out using presentation software. Participants were seated in a sound proof booth in the recording studio at the University of Potsdam in front of a condenser microphone. Participants were familiarized with the task through written and verbal instructions. Each trial consisted of a presentation of the question and its answer on the computer screen. As soon as the sentences were presented, participants heard the prerecorded question, spoken by a male voice. Participants were instructed to speak out the answer displayed on the screen as a response to the question. If the question was answered without any hesitations or false starts, the next trial was presented. If there were hesitations, participants were asked to repeat the answer. Presentation flow was controlled by the experimenter, and participants were allowed to take a break whenever they wanted. The sentences produced by participants were recorded digitally on a computer.

2.3. Participants

18 native speakers of Standard German spoken in the Berlin region participated in the experiment. All were female students at the University of Potsdam. Each speaker was paid or given course credits for participation and took approximately 35 minutes to complete the experiment.

2.4. Data pre-processing and statistical analysis

For this study the first 10 speakers have been selected for annotation and analysis. This resulted in a total of 400 utterances (10 speakers \( \times \) 8 items \( \times \) 5 conditions). The recordings were digitized at a sampling frequency of 44.1 kHz and 16 bit resolution. Data were labeled in Praat [6] by hand at the level of the syllable. Standard segmentation procedures were applied.\(^1\)

For each target word the duration was detected using a Praat script. A repeated measures ANOVA was carried out with duration as the dependent variable. The repeated measures ANOVA disregards the mean of each individual participant (the grand mean of each participant equals to zero), so that there is no additional need of normalizing for the factor speaker.

\(^1\)A neglectable number of the long sentences (0.4 \%) were uttered as two intonation phrases.

![Figure 1: Mean duration pooled across 10 speakers in ms with 95% confidence intervals for the target word Liehner given for wide (All-new), narrow, contrastive focus and prefocus (pre-Given) and postfocal (postGiven) givenness.](image)

![Figure 2: Mean duration pooled across 10 speakers in ms with 95% confidence intervals for the target word Mohn given for wide (All-new), narrow, contrastive focus and prefocus (pre-Given) and postfocal (postGiven) givenness.](image)

### 2.5. Predictions

Based on the above discussed studies, for the factor information structure an increase in duration for focus, and a decrease in duration for givenness is expected. For the factor position, longer durations for earlier target words are expected. And for the factor sentence length shorter duration in longer sentences is expected.

### 3. Results

#### 3.1. Information structure

Figure 1 and 2 present the results for the analysis of duration in different information structure conditions. Mean duration pooled across speakers in ms with 95% confidence intervals is displayed on the x-axis, and the five different conditions are given on the y-axis. The wide focus condition (‘All-new’) is considered to be the baseline condition with a mean duration of 283 ms for Liehner (Fig. 1) and 298 ms for Mohn (Fig. 2).

Comparing the narrow focus condition (‘Narrow’) with the baseline duration increases by 14 ms for Liehner (Fig. 1) and 7 ms for Mohn (Fig. 2). For the contrastive focus condition (‘Contrast’) duration increases by 30 ms for Liehner (Fig. 1) and 22 ms for Mohn (Fig. 2) compared to the baseline.
Comparing givenness conditions with the baseline the two givenness conditions differ in their durational patterns. A prefocally given constituent is on average 22 ms shorter for *Liehner* (Fig. 1) and 33 ms, for *Mohn* (Fig. 2). In contrast, a postfocal constituent is on average 7 ms shorter for *Liehner* (Fig. 1), but 12 ms longer for *Mohn* (Fig. 2). In general, a prostfocal given constituent is longer than a prefocal one, the postfocal given constituent being roughly equivalent in duration to the baseline.

The results which are presented for the disyllabic target word are equivalent to measurements taken only on the first (stressed) syllable of that target word.

### 3.2. Sentence length

Figure 3 and 4 present the results for the analysis of duration for two different sentence length conditions. Mean duration pooled across speakers in ms with 95% confidence intervals is displayed on the x-axis, and the two different conditions of sentence length are given on the y-axis.

Both figures show that the target constituent is longer in shorter sentences. However, this result is statistically only borne out for the monosyllabic item *Mohn* (Fig. 4). The disyllabic target word does not show a significant difference in duration between short and long sentences. Yet, the data show the same trend as for the monosyllabic target word.

The monosyllabic target word is on average 18 ms shorter when uttered in a longer sentence (Fig. 4), while the disyllabic one only about 3 ms (Fig. 3).

### 3.3. Position

Figure 5 and 6 present the results for the analysis of duration for two different position conditions. Mean duration pooled across speakers in ms with 95% confidence intervals is displayed on the x-axis, and the two different conditions of sentence position are given on the y-axis.

With respect to the factor position, the two target words behave different. While for the disyllabic target word no significant difference in duration could be established, the monosyllabic target word differs significantly between these two conditions. The duration of *Mohn* is on average 28 ms shorter when uttered in longer sentences (Fig. 6).

### 4. Discussion and conclusions

The predictions of the present production study are by and large borne out. Based on previous work on duration and focus prominence [1, 4, 8, 9, 10, 11, 15] focused constituents were expected to be longer than compared to the baseline of an all-
new sentence. The data presented here are significantly longer for narrow and for contrastive focus \( (p < .05 \) for both target words). This confirms the prediction of an increase of duration in case of focus prominence.

The raw data suggest that contrastive focused constituents are longer than narrowly focused ones. However, no significant difference between a narrow and a contrastive focus was found. Based on this finding we can conclude that focus as an information structural category has an influence on the durational patterns, yet semantically different focus types \([16]\) do not show categorical differences in the phonetic implementation. From a phonological point of view, thus, we assume a distinction between focus (no matter of narrow or contrastive focus) and no particular focus (all-new sentences).

As for givenness, the data do not show a coherent pattern for prefocus and postfocus givenness. Based on the literature a decrease in duration was expected for given constituents \([11]\). However, this is only true for prefocus given constituents. Postfocus ones are about the same in length as in the baseline condition. This distinction has not been reported yet, and the reason for relatively long postfocal constituents is by no means clear.

Apart from the effect of information structure two further effects on the duration of constituents have been investigated, the effect of the length of the sentence and, of position in the sentence on the target word. The former one has to our knowledge not been investigated so far, though it is expected that constituents in a longer sentence overall decrease in duration. The present data proves this effect, yet only for the monosyllabic target word. For the disyllabic target word, only a trend in the same direction can be observed. From the present data set it is not clear to what extend the length of a sentence influences the durational patterns.

As for the effect of position of the target word in the sentence it has been claimed that the duration of constituents is longer if they occur early in the sentence as compared to late \([11, 12]\). The data of the present study has confirmed this claim only partly. The monosyllabic target word was affected by position while the disyllabic target was not. It is unclear why the effect of position only affects the monosyllabic target word in this study. The only obvious difference between the carrier sentences of the two target words is their number of syllables. The target word Liehner was embedded in a sentence containing 20 syllables, while Mohn was embedded in a sentence containing 13 syllables. It may be the case that the higher number of syllables, i.e. longer overall sentence duration, obscured the durational decrease of the target word. Another explanation would be that phrasing longer sentences into two intonation phrases might cause this effect; however, this cannot be confirmed by our data since only seven out 200 long sentences have been uttered in two IPs.

Overall, this study has confirmed the effects of information structure on duration while the effects of sentence length and position of a target word in the sentence need further investigation. However, this study has provided quantitative data that can be used as a starting point for manipulating speech data in order to perceptually test the role of duration for the recognition of focus prominence.

Acknowledgments
The research presented here was funded by the Deutsche Forschungsgemeinschaft (DFG), through the project D5 of the research centre SFB 632 on Information Structure at the University of Potsdam. Thanks to Anja Gollrad, Sophia Jähningen, Gerrit Kenntner and Bernadett Smolibocki for assistance.

5. References


PRODUCTION AND PERCEPTION OF CONTRAST IN GERMAN

Frank Kügler & Anja Gollrad

Department of Linguistics, University of Potsdam
kuegler@uni-potsdam.de, agollrad@googlemail.com

ABSTRACT
In German intonational phonology two different approaches to model nuclear falling accents exist (cf. [7], [8] vs. GToBI [11]). While [7] and [8] assume only left-headed pitch accents, the GToBI system allows for left- and right-headedness. Consequently, in the former approach one simple falling accent (H*L) is represented in the tonal grammar while GToBI distinguishes between two falling accents, a simple falling and a rising-falling one (H*L- and L+H*L-). The present production and perception studies aim at providing phonetically based evidence in order to argue for one of the two phonological approaches of representing falling accents in German.

Keywords: falling accents, focus, contrast, tonal perception, German

1. INTRODUCTION
The concept of contrastiveness has received much attention in psycholinguistic research [e.g., 4, 7, 13, 14, 18]. However, the phonetic and/or phonological nature of accents realized under a contrastive context compared to a neutral accentuation is still a matter of debate, i.e., there is a diversity of interpretations with respect to the phonological category falling accent types belong to in German.

According to GToBI there is a difference between two types of falling accents, which is accompanied by a difference in intonational function, i.e. a difference in focus structure [10]. According to [7], however, there is no tonal distinction between narrow and wide focus in German. Both [7] and [8] claim that in case of falling accents the pitch shape prior to the pitch peak is phonetic in nature, thus no further low leading tone is necessary to present falling accents. [8] proved this claim by means of a discrimination test showing that grouping of her data according to two different tonal categories failed. All studies provide poor evidence in favour of their claims: [7] argues phonologically not providing detailed phonetic evidence, [8] argues perceptually not referring to information structure, thus not considering the distinction assumed by GToBI, and [11] argue functionally without further phonetic evidence except for inter-transcriber consistency [12]. The present research is approaching the debatable issue by a combined production and perception study to examine how falling accents are represented in the tonal grammar of German.

2. EXPERIMENTS
In speech production the realization of wide vs. narrow focused sentences were compared investigating the phonetics of falling accents in nuclear position in German, and how the realization can be represented phonologically.

2.1. Production

2.1.1. Speech materials
The experimental sentences contain a SAuxOV word order with target words embedded as objects in non sentence final position in order to avoid any intonational phrase boundary effects. The following factors were manipulated:
• The number of accents in the sentence: one and two accents.
• The number of syllables of the target word varied between one (Wal ‘whale’), two (Roman ‘novel’), and three (Admiral ‘admiral’), all with ultima word stress.
• The length of the sentence: sentences were gradually lengthened by adding one of the two adverbials (gestern ‘yesterday’), and (glücklicherweise ‘luckily’) or a combination of both prior to the target word to increase the inter-accentual distance (between a sentence initial, and an accent on the target word. We expected that a larger inter-accentual distance would increase the chance that speakers realise two single peak accents instead of a hat pattern, which is a frequent pattern in German.
(1a) illustrates an example of a wide focus target sentence, (1b) illustrates a target sentence realized under contrastive focus. In both sentences, the target word is monosyllabic.

(1a) Erzähl mir bitte, was passiert ist.
‘Please tell me what happened.’
Martin hat den Wal gesehen.
‘Martin has seen the whale.’

(1b) Hat Martin den Frosch gesehen?
‘Has Martin seen the frog?’
Nein, Martin hat den Wal gesehen.
‘No, Martin has seen the whale.’

The experimental sentences are highly sonorant to allow for a maximally accurate f0 analysis. Sentences were interspersed with numerous fillers and fed into DMDX [5]. The experimental sentences were pseudo-randomized for each subject so that sentences of the same condition did not appear adjacently and corresponding sentences had a maximal distance.

2.1.2. Speakers

8 speakers participated in the experiment. All were female undergraduate students at the University in Potsdam. All were native speakers of standard German spoken in the Berlin-Brandenburg region and reported no speech or hearing impairment. They either received course credit or were paid for participation.

2.1.3. Recording procedure

For each sentence, a context question eliciting wide focus (1a) and contrastive focus (1b), spoken by a male voice, had been previously recorded. The contexts were presented together with a target sentence both visually on screen and auditorily over headphones. Speakers were asked to read and listen to the context and then to say the answer displayed on the screen as a response to the question. Subjects were familiarized with the task through written and verbal instructions. In case of hesitations or false starts, participants were asked to repeat the sentence. Recordings took place in a sound-proof chamber equipped with an AT4033a audio-technica studio microphone, using a C-Media Wave soundcard at a sampling rate of 44.1 kHz with 16 bit resolution. Presentation flow was controlled by the experimenter, and participants were allowed to take a break at any point.

A total of 486 target sentences had been recorded (discarding 11% due to creaky voice and mispronunciations). The resulting 432 sentences were submitted to an extensive perceptual inspection which revealed four distinct phonological contours of the target sentences realized by the speakers. The following patterns were categorized: (a) 299 non-downstepped nuclear falling accents, (b) 25 downstepped nuclear falling accents, (c) 36 hat patterns, and (d) 72 other types of nuclear accents, such as early falls. The phonetic examination of falling accents was based on group (a). Since GToBI assumes rising-falling accents we were particularly interested in contours that contain a nuclear peak accent, hence hat patterns were excluded as well as downstepped accents. The distribution over the broad- and contrastive focus cases of the 133 tonal accents, excluded from further analysis, was as follows: 79 broad focus cases and 54 contrastive focus cases (for similar figures see [9]).

The experimental sentences were hand-annotated and subjected to phonetic analysis using Praat software [3]. The following phonetic measurements were conducted: (1) the pitch peak of the target words in Hertz (Hz) and the corresponding time of the peak (tH), (2) a low turning point in pitch prior to the peak in Hz (l) and the corresponding time of the low point (tl), (3) the beginning and the end of the accented syllable, (4) the excursion between the low pitch point and the peak, (5) the velocity of the rise, and (6) the alignment of the low turning point in relation to the accented syllable.

![Figure 1: Phonetic measurements of the target word](image-url)

Pitch analysis was conducted using a Hanning window of 0.4 seconds length with a default 10 ms analysis frame. The pitch contour was smoothed using the Praat [3] smoothing algorithm (frequency band 10 Hz) to diminish microprosodic perturbations.
2.1.4. Results

We fitted a multilevel model [2] using crossed random factors subject and item, and focus condition (wide focus (WF), contrastive focus (CF)) as fixed factor. The acoustic analysis revealed no significant difference between the focus conditions of the factor alignment of the low turning point. The analysis revealed only a marginal significant difference between the focus conditions for the f0 of the low turning point (WF: 188.07 Hz, CF: 182.45 Hz, t=2.73), for the f0 of the pitch peak (WF: 232.87 Hz, CF: 233.90 Hz, t=2.93) and for the alignment of the peak (WF: -0.055, CF: -0.065, t=2.08). The pitch excursion (WF: 44.79, CF: 51.45, t=-4.592) and the velocity of the rise (WF: 254.28, CF: 297.80, t=-5.19) yield significant differences between the two focus conditions.

The small differences between the two focus conditions with respect to (a) the alignment and (b) the Hz-values of the low turning point and the peak raises the question if and to what extent listeners would perceive a difference of the targets words as a function of focus.

2.2. Perception

A semantic congruency task [15] was conducted to investigate whether German listeners use the small phonetic differences of (a) the low target prior to the high tone and (b) of the high tone itself in the nuclear falling accent as a primary perceptual cue to distinguish accents in contrastive focus from accents in wide focus. Semantic congruity tests have been successfully used to explore the perception of intonation contrasts [16, 17]. The test allows us to evaluate the degree of perceived appropriateness of target intonation patterns to different pragmatic contexts.

2.2.1. Material

Stimulus materials for the perception experiment were taken from one of the speakers of the production study who produced the most prominent difference from the mean value of the low turning point in the two focus conditions. Context questions and target sentences were concatenated to one sound chain at a sampling frequency of 48000 Hz and were scaled at an intensity of 70db. The perception experiment consisted of 12 target sentences where intonation was coherent with the pragmatic context (6 WF-WF dialogs, 6 CF-CF dialogs) and 12 target sentences where intonation was not coherent with the pragmatic context (6 CF-WF dialogs, 6 WF-CF dialogs). Each dialog was presented 3 times which resulted in a total of 72 dialogs. The stimuli were auditorily presented over headphones with the MFC Praat software [3].

16 listeners were asked to listen to each dialog carefully and then evaluate whether they regard the intonation of the target sentence as “congruent” (by clicking on the “congruent box” visible on the screen) or as “incongruent” (by clicking on the “incongruent box”). After written and verbal instructions, a test run of 3 dialogs was carried out before the experiment started. The experiment lasted approximately 20 minutes.

2.2.2. Results

Figure 2: Rate of congruous responses to all dialog types, separated by WF-context (left) and CF-context (right).

Figure 2 displays the rate of congruent responses to all dialog types, separated into WF-context (left bars) and CF-context (right bars). The results revealed that the appropriateness of the target intonation pattern to a context were rated higher for congruent (WF-WF and CF-CF) than for incongruent dialog types (CF-WF and WF-CF). Linear mixed-effects models [2] revealed that the difference of the rating response between the two conditions (congruous and incongruous) were significant for the WF-context dialogs (WF-context: t=11.75) as well as for the CF-context dialogs (CF-context: t=11.25). These results indicate that listeners are extremely sensitive to congruous and incongruous intonation of a context-target sequence.
3. CONCLUSION

A production experiment had been carried out testing the influence of wide focus and contrastive focus contexts on the realization of nuclear fallings accents in German. The analysis revealed that the phonetic difference of the target words is only marginal with respect to (a) the low turning point prior to the peak, (b) the f0-value of the peak itself and (c) the alignment of the peak in both focus conditions.

The fact that contrastive focus raises nuclear H* accents in German is well-known, e.g. [6]. If there are two distinct falling accents in German [11], we would expect a difference in phonetic realization regarding the low turning point prior to the high accent. This is indeed the case for scaling. The low turning point is realized lower in contrastive focus than in wide focus. However, there is no significant difference in alignment of this low turning point. Given the marginal acoustic differences the question remains whether listeners perceive a difference between the wide and contrastive focus elicitation at all.

A semantic congruency experiment using congruous (WF-WF and CF-CF) and incongruous dialogs (WF-CF and CF-WF) was conducted aiming to investigate whether listeners are able to perceive the phonetic difference as a function of focus. Interestingly, the results of the perception study show that listeners are able to distinguish between a congruous and an incongruous dialog, in the WF-context as well as in the CF-context condition (see Fig. 2). In sum, the accents realized under contrastive focus which have a lower f0 turning point and a later as well as a higher f0 peak compared to wide focus accents can be distinguished from each other perceptually.

In order to answer the question of tonal representation, the question remains whether the lower scaling of the low turning point causes the perceptual impression of contrastive focus, or whether higher scaling of H* accents is a sufficient phonetic cue. The fact that listeners differentiate between marginally distinct intonation contours as recorded in our production data allows for a further perception study manipulating the individual cues.

Acknowledgements: This research was supported by the DFG-grant (KU 2323/1-2) to the project “Prosody in Parsing” at Potsdam University, principle investigators: Frank Kügler, Caroline Féry and Shravan Vasishth. We are grateful to Tobias Günther who provided considerable technical assistance.

4. REFERENCES

Section II
Hindi
Focus, Word Order and Intonation in Hindi

UMESH PATIL, GERRIT KENTNER, ANJA GOLLRAD, FRANK KÜGLER,
CAROLINE FÉRY, SHRavan VASISHTH, University of Potsdam

Received November 27, 2007; Revised July 8, 2008

ABSTRACT

A production study is presented that investigates the effects of word order and information structural context on the prosodic realization of declarative sentences in Hindi. Previous work on Hindi intonation has shown that: (i) non-final content words bear rising pitch accents (Moore 1965, Dyrud 2001, Nair 1999); (ii) focused constituents show greater pitch excursion and longer duration and that post-focal material undergoes pitch range reduction (Moore 1965, Harnsberger 1994, Harnsberger and Judge 1996); and (iii) focused constituents may be followed by a phrase break (Moore 1965). By means of a controlled experiment, we investigated the effect of focus in relation to word order variation using 1200 utterances produced by 20 speakers. Fundamental frequency (F0) and duration of constituents were measured in Subject-Object-Verb (SOV) and Object-Subject-Verb (OSV) sentences in different information structural conditions (wide focus, subject focus and object focus). The analyses indicate that (i) regardless of word order and focus, the constituents are in a strict downstep relationship; (ii) focus is mainly characterized by post-focal pitch range reduction rather than pitch raising of the element in focus; (iii) given expressions that occur pre-focally appear to undergo no reduction; (iv) pitch excursion and duration of the constituents is higher in OSV compared to SOV sentences. A phonological analysis suggests that focus affects pitch scaling and that word order influences prosodic phrasing of the constituents.

1 Introduction

Prosody is an integral component of language, and so it is only natural that online sentence comprehension and production critically depend on the structuring cues provided by prosody. The role of prosody in comprehension and production seems to be especially relevant for languages such as German and Hindi that involve relatively free word order; word order flexibility increases the number of options available for expressing information structure, significantly complicating the means by which the incoming signal can be decoded. Any extra-syntactic cue, such as prosody, would a priori be expected to provide an important cue for facilitating comprehension.

European languages such as English and German (e.g., Gussenhoven 2008, Ladd 1996, Selkirk 2007, Truckenbrodt 1995, Féry and Küglar 2008) have received a great deal of attention concerning the prosodic marking of information structure. However, not much is known about Hindi in this respect. In an attempt to fill this gap, we investigate the effect of word order and intonation on prosody. We carried out a production study of Delhi Hindi (20 participants) which showed that the intonation of Hindi and its interaction with focus and word order differs from well-studied intonational languages such as English and German. First, the primary prosodic cue accompanying focus...
on a constituent is post-focal pitch range reduction, rather than the raising of F0 as observed in intonational languages such as English; this is consistent with previous research on Hindi intonation (Moore 1965, Harnsberger 1994). In pitch range reduction (or compression), the voice register available for realizing the melody of tones is reduced as compared to the one the speakers have at their disposal when starting the sentence. High tones are lower, or not realized at all, depending on the amount of register compression. Second, in sentence-initial focus, canonical (SOV) word order shows a greater post-focal pitch range compression than non-canonical (OSV) order. Third, when sentence-medial elements are focused, (i) the sentence-initial object in non-canonical (OSV) order has a higher F0 peak as well as a larger F0 range than the sentence-initial subject in canonical (SOV) order, (ii) the duration of the medial (focused) element is longer in the non-canonical order compared to canonical order, and (iii) no evidence is found of pre-focal pitch range compression for given (previously mentioned) elements.

We propose an analysis of phrasing in Hindi according to which each content word in Hindi is phrased separately as a prosodic-phrase (p-phrase). Each p-phrase receives a low tone, which we analyze as a pitch accent, and a high phrase boundary associated with the right edge of the prosodic word. Regarding pitch range effects, tones are scaled relative to abstract reference lines, and we assume the all-new sentence pattern to represent the neutral baseline. Focus, then, compresses the post-focal register in Hindi.

2 Background

Hindi belongs to the Indo-European branch of languages and is an Indo-Iranian language (Kachru 1987) genetically related to European intonation languages such as English, German and Russian. It is natively spoken by approximately 366 million speakers (source: Ethnologue, www.ethnologue.com), mainly in the central and northern part of India, but also in Bangladesh, Nepal, the United Kingdom and many other countries. In addition, Hindi is also used as a second language or a lingua franca by many Indians in non-Hindi speaking regions. The syntax of Hindi and Urdu (which is spoken in Pakistan as well as India) are virtually identical, although the scripts and choice of content words differ. We use the term ‘Hindi’ in this paper but the conclusions presented are expected to extend to Urdu as well.

2.1 Hindi intonation

The work of Nair (2001) and Dyrud (2001) suggests that Hindi has lexical stress, in the sense that every word has a designated syllable on which prominence is realized (see Hayes 1995, Moore 1965, Ohala 1986, who notice contradictions on this view). Nair (1999) and Dyrud (2001) find acoustic correlates of prominent syllables, like higher pitch and longer duration. Position of lexical stress is dependent on syllable weight, the heavier syllables attracting stress first. Hussain (1997) shows how the position of stress can be predicted by syllable weight (but see Ohala 1986, who finds differences in stress position depending on whether a word is uttered in isolation or in a sentence).

All researchers on Hindi intonation appear to agree that each content word except the final one is associated with a rising contour.1 According to Moore (1965, 68), every p-phrase (called ‘foot’ by Moore) contains a pitch accent, where this prosodic domain is defined as “one to several syllables in length, which normally is uttered with a continuously rising pitch from beginning to end”. Harnsberger (1994) makes a similar observation, and proposes that the low part of the rising contour is a low pitch accent, annotated as L* in an autosegmental-metrical notation system (Pierrehumbert 1980). The high part of the rising contour is either a trailing tone H−2, or a boundary tone Hp. The subscript ‘P’ represents a phrase boundary smaller than the intonation phrase (see Hayes and Lahiri 1991, for this annotation convention). In the analysis presented below, we treat the high part as a boundary tone.

---

1Moore’s data suggests that wh-questions are realised with a different intonation pattern, though Moore himself does not emphasize this fact and a systematic analysis has yet to be done.

2A trailing tone is the part of a bitonal pitch accent that follows the starred tone.
Moore’s account of Hindi intonation comprises an analysis of pitch, intensity and duration for three prosodic phenomena: the expression of emphasis (or focus), the expression of speaker attitudes, and phrasing (Moore 1965, 62). Although Moore distinguishes different melodic contours on a single syllable, i.e. level, rise, fall, rise-fall, fall-rise (Moore 1965, 65), the underlying pitch accent is considered a rising one. Deviations from this underlying pattern on the surface are the result of speaker’s attitude, or, alternatively, the result of a pure phonetic effect, namely tonal transition from a very high rising pitch to the following low tone (1965, 68, 75).

According to Moore, focus has a phrasing effect. In the terminology used here and which is also used used by Lahiri & Hayes for the closely related language Bengali (Hayes and Lahiri 1991), focus has the effect of inserting a p-phrase boundary at the left edge of the focused phrase. The same effect of focus has also been observed in Harnsberger (1994). According to Hayes & Lahiri, a focused phrase also causes dephrasing of the postfocal material. Like Hindi, Bengali is a head-final language (SOV), and the formation of p-phrases is primarily based on syntax (see also Lahiri and Fitzpatrick-Cole 1999). Every maximal constituent is a p-phrase, and the verb often forms its own p-phrase, obligatorily so when it is focused. Evidence for p-phrases comes not only from the tonal structure, but also from segmental processes like /r/ assimilation and voicing assimilation, which only take place inside of p-phrases. The absence of these processes correlates with p-phrase boundaries. Some variations in phrasing occur as a consequence of speed, style and givenness, but these variations are also subject to syntactic constraints. It is not the case that all kinds of restructuring are allowed. Hayes & Lahiri also show that the phrase construction rule is cyclic and recursive. In contrast to Moore and Harnsberger, who assume two levels of prosodic phrasing below the intonation phrase, Hayes & Lahiri only allow p-phrases, which are cyclic and recursive. This is also the view adopted in this paper, though our experimental sentences do not make use of recursive p-phrases.

Beside the phrasing effect just discussed, focus has been claimed to have two additional prosodic effects (see Harnsberger 1994, 1999, Moore 1965, Dyrud 2001, among others). First, the rising pitch pattern may show a higher excursion, a greater intensity and longer duration. Second, post-focally the pitch range may be compressed or even completely flat and deaccented (Harnsberger and Judge 1996), although rising pitch accents are still realized in compressed pitch range.

There is a striking similarity between the intonation of Hindi and detailed descriptions of other South-Asian languages like Bengali and Tamil: Hayes and Lahiri (1991) and Khan (2007) assume several similar prosodic properties for Bengali, such as lexical stress, pitch accents associated with stressed syllables, and intonational phrasing on at least two levels. The Dravidian language Tamil, although genetically unrelated, also shows similarities with Hindi (Keane 2007a,b); this could be a consequence of language contact. According to the sentence-based typology of intonation systems recently developed by Féry and Fanselow (2008), Indo-Aryan and Dravidian languages share the property of being phrase languages, that is languages in which tones assigned at the level of the p-phrase play a crucial role for the intonational pattern. These languages differ from intonation languages like English for instance, which freely assign different kinds of pitch accents on words.

2.2 Information structure and Hindi syntax

Hindi is a head-final (Subject-Object-Verb) language, with relatively free word order. Constituents may be scrambled to express different information structural configurations, or for stylistic reasons. The first syntactic constituent in a sentence is usually the aboutness topic (Gambhir 1981, Butt and King 1996), which may under certain conditions be marked by the clitic -to, similar in some respects to Japanese -wa (Kuno 1981, Kidwai 2000). The term ‘aboutness topic’ is understood here as a referent which the remainder of the sentence is about, possibly contrasting with other referents, and followed by a focused constituent (see Reinhart 1981, Jacobs 2001, among others).

In Hindi, a focused constituent typically occupies the immediately preverbal position, and wh-

\footnote{Note, however, that according to Moore (1965) a focus needs not necessarily be realized by means of all correlates but any combination of one or two of these may suffice to phonetically express focus. Thus, F0 as a correlate may be absent.}
markers also tend to occur preverbally (Kidwai 2000, 116). Nominal clitics can serve to mark focus (similar to English focus particles like ‘only’, ‘even’ or ‘also’, Sharma (2003)). However, focused constituents need not be morphologically marked. In this paper, ‘focus’ is used rather traditionally as the part of the sentence which introduces alternatives (Rooth 1985, 1992). The term ‘focus’ is applied to constituents which are informationally more important than other backgrounded parts of the same sentence. In the general case, an all-new sentence does not trigger a set of alternatives, though the possibility of focusing a whole sentence cannot be excluded in principle. Below, we call an all-new sentence a ‘wide-focused’ sentence. According to Butt and King (1996), in situ focusing of a phrase in Hindi is possible with multiple foci and results in contrastive focus readings. Kidwai (2000, 114-137) presents detailed arguments that focus is responsible for scrambling operations such as preposing (as XP adjunction operations).

Butt and King also provide evidence that background information occurs postverbally, and completive information – which is information of secondary importance to the information structure of the discourse – occurs in the preverbal region preceding the focus position.

2.3 Aim of the present study

Thus, although much is known or hypothesized about word order and focus in Hindi, very few controlled experimental studies exist that explore the interaction with prosody. In order to remedy this situation, we took this previous work as a theoretical starting point and designed a production experiment that investigates the intonational realization of focus and its interaction with different word orders.

3 Production experiment

3.1 Method

3.1.1 Design and Materials

The experiment involved a 3x2 factorial design with two factors: focus (subject, object and wide focus) and word order (SOV and OSV).

Each trial consisted of a question-answer pair: a question and a response to the question. The question set up either a subject, object or wide focus for the response utterance; see examples (1), (2), (3). In (1-b) and (1-c) the question involves subject focus, in (2-b) and (2-c) object focus, and in (3) wide focus. In the examples, a bracketed segment with a subscripted F stands for the focused element relative to the preceding question.

(1) Subject question

a. kis ne davaaii ko khariidaa?
   who ERG medicine ACC buy.PAST?
   ‘Who bought the medicine?’

b. [graahak ne]F davaaii ko khariidaa
customer ERG medicine ACC buy.PAST
   ‘(The) customer bought the medicine.’

c. davaaii ko [graahak ne]F khariidaa
   medicine ACC customer ERG buy.PAST
   ‘(The) customer bought the medicine.’

(2) Object question

a. graahak ne kyaa khariidaa?
customer ERG what buy.PAST?
   ‘What did the customer buy?’

---

4 Only two word orders (SOV and OSV) are considered here in order to keep the number of experiment conditions tractable, and because not all word orders are possible in Hindi and the constraints on word order variation are far from clear.
(3) Wide focus question
   a. kyaa luuaa?
      'What happened?'
   b. [graahak ne davaaa ki ko khariidaa]F
   c. [davaaa ki ko graahak ne khariidaa]F

The questions were always in canonical (SOV) word order, and the answers always contained a transitive verb and two arguments, with subject arguments in ergative case and objects in accusative case. Subjects were nouns referring to humans and objects referred to inanimate referents. Past tense and perfective aspect was used in all sentences. Half of the subject and object nouns were bisyllabic with initial stress and the other half were trisyllabic with stress on the second syllable. The complete set of target items is shown in the Appendix.

A note on terminology: Since at most one argument (subject or object) is focused in the question utterance, the non-focused argument will be designated as given in the response utterance (previously mentioned in the question; Allerton 1978, Lambrecht 1994). For example, in the subject question (1), davaaa, ‘medicine’, is mentioned, and so in the response utterances (1-b) and (1-c) the referent denoted by davaaa is given. By contrast, in the wide focus condition (3), since none of the arguments are mentioned, in the corresponding response utterances (3-b) and (3-c) none of the noun phrases refer to given elements. This distinction between the focused and given element becomes relevant when the results of the experiment are presented.

A total of 18 unique question-answer pairs were constructed and each pair was realized in the 6 conditions, resulting in 18×6=108 sentences per speaker. All the 108 sentence-pairs were presented to each speaker in a pseudo-randomized manner; items from four other unrelated experiments were interspersed as fillers. Four pseudo-randomized lists were prepared to minimize order effects.

The questions were recorded in a speech recording laboratory in the University of Potsdam in preparation for presentation of stimuli to participants; the presentation procedure is described below.

3.1.2 Participants
30 native speakers of Hindi participated in the experiment. All were female students at the University of Delhi, India and were residents of Delhi and surrounding areas. Each speaker was paid 150 Indian Rupees for participation and took approximately 45 minutes to complete the experiment.

3.1.3 Procedure
The experiment was carried out using presentation software. First, participants were equipped with a set of headphones and a microphone head-set, and familiarized with the task through written and verbal instructions, followed by two practice trials. Each trial consisted of a presentation of the question and its answer on the computer screen, written in Devanagari. Participants heard the pre-recorded question over headphones, spoken by a male voice. At the same time the target sentence was presented on the screen. Participants were instructed to speak out the answer displayed on the screen as a response to the question. If the question was answered without any hesitations or false starts, the next trial was presented. If there were hesitations, participants were asked to repeat the answer. A total of 48 items (4%) had to be repeated because of false starts or hesitations. Presentation flow was controlled by the experimenter, and participants were allowed to take a break whenever they wanted. The sentences produced by participants were recorded at the University of Delhi on a DAT tape recorder.

3.1.4 Data pre-processing and statistical analysis
Due to limited resources we analyzed a subset of the data. Of the 18 items, five items from each syllabic and stress pattern, i.e., a total of 10 items, were selected for annotation and analysis (the
**Dependent variable** | **Where measured**
---|---
F0-maximum (Hz) | at the right edge of preverbal constituents
(F0-minimum (Hz) at the right edge of preverbal constituents) (This is where H boundary tones are expected.)
F0-range (Hz) | preverbal constituents
Duration (ms) | preverbal constituents

**Table 1** Dependent variables.

first five items of each syllabic pattern) Of the 30 speakers, utterances from the last 20 speakers was used in the data analysis. This resulted in a total of 1200 utterances (20 speakers × 10 items × 6 conditions). The above criteria for subsetting the data were decided upon arbitrarily.

The recordings were redigitized from DAT at a sampling frequency of 44.1 kHz and 16 bit resolution. Data were labeled by hand at the level of the constituent, as shown in (4). The vertical lines mark constituent boundaries.

(4) | graahak ne | davaaii ko | khariidaa |
---|---|---|---|
 customer | ERG | medicine | ACC | buy.PAST
‘(The) customer bought the medicine.’

The pitch analysis was conducted using a Hanning window of 0.4 seconds length with a default 10 ms analysis frame. The pitch contour was smoothed using the Praat (Boersma and Weenink 2005) smoothing algorithm (frequency band 10 Hz) to diminish microprosodic perturbations. Stylized pitch tracks were calculated. For this purpose, each constituent in (4) was divided into five equal intervals, and the mean pitch was aggregated over the 20 speakers and 10 sentences for each interval. The resultant values were interpolated separately for each condition.

For each constituent in (4), the maximum F0, the minimum F0 and the duration were detected using a Praat script. In the second constituent, only those F0 maxima were measured that followed the F0 minimum in that constituent; this was done in order to exclude maxima due to transitions from preceding H tones. The maximum after the low tonal target represents the high tone in the LH gesture. Based on the measurements of F0-maximum and F0-minimum the F0-range was calculated (F0-max minus F0-min).

The statistical analysis relied on three dependent variables, F0-maximum, F0-range, and duration; the loci of these measurements are shown in Table 1. All dependent measures were log-transformed to meet the assumption of the regression model.

A multilevel model (Gelman and Hill 2007, Bates and Sarkar 2007, Pinheiro and Bates 2000) was fit, using crossed random factors speaker and item, and focus status of constituent (wide focus, narrow focus, given), and word order of sentence (SOV vs. OSV) as fixed factors.

### 3.2 Results and Discussion
#### 3.2.1 Effect of Focus

The contours in Figures 1–3 show time-normalized mean pitch tracks for each focus condition averaged over all 20 speakers. The contours show rising tonal patterns on the non-final constituents and falling patterns on the final verb. Table 2 in the Appendix summarizes the results of the statistical analyses.

As shown in Figure 1, for SOV structures, in the subject, object and wide focus conditions a rising pitch gesture is realized on both pre-verbal constituents. Object focus and wide focus do not show a significant difference, but subject focus is realized with a significantly higher F0 excursion compared to the wide focus counterpart (although the magnitude of the difference is small). In the subject focus condition, the rising gesture on the object is realized in a clearly lower and compressed range compared to the other conditions.

A similar pattern is seen in OSV structures. Here, the pitch tracks of subject focus and wide focus are nearly identical. This absence of a difference between the two types of focus may due to
FIGURE 1 Time-normalized pitch tracks, based on five measuring points per constituent, showing the mean across all speakers. The upper plot shows SOV order and the lower plot OSV order. The comparisons of interest in each plot are subject focus (dotted line) and object focus (dashed line) with respect to wide focus (solid line).
the fact that in OSV word order, the subject is in the default preverbal focus position (Section 2.2) in both the conditions. In OSV structures, the realization of (sentence-initial) object focus shows a divergent pattern, although not on the focused constituent itself. Only post-focal compression is visible: the rising pitch gesture in the post-focal constituent is significantly lower than the one for wide focus.

Focus thus induces post-focal compression of the pitch range, which confirms the results of Moore (1965), and Harnsberger and Judge (1996). However, the expected effect of greater pitch excursion on the focused constituent itself is only borne out for initial focus in SOV sentences but not for medial focus or OSV sentences.

We now turn to the results of the statistical analyses on the three dependent variables. Subject focus in SOV order had a higher F0-maximum (t=4.26), a greater pitch range (t=4.97), and longer duration (t=2.62) on the focused subject, compared to the wide focus baseline. Compared to the baseline, the given object showed a significantly lower F0-maximum (t=−9.06), a smaller F0-range (t=−9.94), and shorter duration (t=−6.24).

Although no effect of focus was found on the object in OSV sentences, post-focal compression on the medial constituent was seen when the initial object was in focus (lower F0-maximum (t=−8.23), lower pitch range (t=−6.7) and shorter duration (t=−3.62) compared to the wide focus baseline) much as in SOV order. When the medial subject was focused, it had a slightly but significantly higher F0 range (t=2.34) compared to the baseline. F0-maximum and duration did not yield significant effects here.

Pre-focally given constituents do not show any clear difference compared to wide focus baseline: an initial given subject is nearly identical in F0 maximum (initial subject and baseline: 274 Hz), F0 range (initial subject 58 Hz, baseline 59 Hz), and duration (initial subject 556 ms, baseline 558 ms). The same result holds for an initial given object (F0-max: initial object 276 Hz, baseline 279 Hz; F0-range: initial object 67 Hz, baseline 68 Hz; and duration: initial object 551 ms, baseline 558 ms).

### 3.2.2 Effects of word order

**Wide focus.** Figure 2 shows SOV and OSV word orders in the wide focus condition, and Table 3 in the Appendix summarizes the results of the statistical analyses.

Almost no difference is seen in the time-normalized pitch tracks. The marked word order (OSV) is on average slightly lower in the rising part of the first constituent, higher on its peak, and it is slightly higher on both the rising part and on the peak of the second constituent. Regarding the dependent variables, the F0-range on the initial constituent, but not the F0-maximum, is significantly larger.

![Pitchtrack (Wide Focus)](image)
Figure 3 Time-normalized pitch tracks based on five measuring points per constituent averaged across all speakers for SOV (solid line) and OSV (dotted line) word order; in the upper plot the first constituent, and in the lower plot the second constituent is in focus.

(9 Hz) in OSV than in SOV, \( t = 4.59 \). This suggests that the low pitch accent is realized lower in OSV order. Additionally, the duration of the preverbal constituent is on average 17 ms longer in OSV compared to SOV. This difference is significant \( (t = 2.51) \).

**First or second constituent in focus.** We turn next to the word order comparison when either the first (Figure 3a) or second constituent is in focus (Figure 3b). In sentence-initial focus (Figure 3a), no difference was found in the realization of the focused constituent, but in the post-focal constituent the amount of post-focal compression is larger for the unmarked word order. In other words, the post-focal pitch range is higher for OSV word order. When the first constituent is the focus, the medial subject (OSV) displays a higher F0-maximum \( (t = 4.55) \), a greater F0-range \( (t = 2.65) \), and longer duration \( (t = 3.76) \) than the medial object (SOV). Although the contour plot (Figure 3a) suggests an earlier and steeper rise in pitch on the first constituent for SOV structures, as compared to OSV sentences, no significant differences in the dependent variables were found on the initial focused constituents.

In case of second-constituent focus, a small difference appears on the pre-focal given constituent. The marked word order (OSV) shows a higher F0 peak on the initial given object. The F0-maximum \( (t = 3.44) \), and correspondingly, the F0-range \( (t = 4.77) \) on the initial constituent is significantly higher in OSV, as compared to SOV structures. The duration of the medial (focused) constituent itself is 20 ms longer in OSV than SOV sequences \( (t = 3.56) \). In the duration measure, the same pattern is
seen for these constituents when they are given (22 ms, t=3.76), i.e., when the focus is on the first constituent.

3.3 Summary of the results

The results can be summarized as follows. Pitch contours of SOV and OSV sentences display the same basic pattern: both arguments have a rising tonal structure, and the verb has a falling structure. This result is in line with previously observed pitch patterns in Hindi (Moore 1965, Harnsberger 1994). The highest part of all three constituents are in a very clear downstep relationship to each other.\(^5\)

With the exception of initial subject focus in SOV structures, focus was not found to affect the pitch excursion and duration of the focused elements. However, it does affect the post-focal constituent, when the initial element is in focus: in both word orders, the medial, post-focal constituent has a lower F0-maximum, a smaller F0-range and duration is shorter than in the baseline wide focus condition. When focus is on the second, preverbal constituent, no reliable difference was found compared to the wide focus baseline in F0-maximum, F0-range and duration on either of the preverbal constituents. Thus, we found no evidence of pre-focal compression due to givenness, and no raising of F0 as a consequence of focus as reported for many intonation languages (e.g., Bartels and Kingston 1994, Baumann et al. 2006, Cooper et al. 1985, Féry and Kügler 2008). The absence of any prosodic effect when the preverbal constituent is focused might be due to the fact that the preverbal position is the syntactic default position for focus (Kidwai 2000). Prosodic marking of focus in this position might therefore be redundant.

Word order has an effect on prosody: significant differences were found between SOV and OSV word orders, appearing most clearly on the given constituents. First, in sentence-initial focus, the amount of post-focal compression was larger in SOV than in OSV sentences, as reflected by a lower F0-maximum, a smaller F0 range and shorter duration of the medial constituent in SOV sentences. Moreover, in sentences with focus on the second constituent, the F0-maximum, as well as the F0-range on the initial given constituent are slightly but significantly higher in OSV, as compared to SOV structures. Third, the duration of the medial focus constituent is longer in OSV than SOV sequences.

4 A phonological interpretation

A phonological analysis (including phrasing and pitch scaling) is presented next, based on the production data. The prosodic phrasing of the experimental sentences is a direct consequence of syntactic structure and is thus very simple, as the syntactic structure of the sentences investigated is quite simple and all constituents are phrased individually. The tonal realization is, on the one hand, dependent on the syntactic structure (downstep pattern), as well as finality or non-finality of the smaller prosodic phrases in a larger intonation phrase, and, on the other hand, the result of information structure. The variation observed in pitch scaling comes from information structure and is discussed in more detail below.

4.1 Phrasing

The general pattern of intonation described by Moore (1965), Harnsberger (1994, 1999) and others was confirmed in our data. Every content word (here every argument) except for the final one (the verb) has a rising contour, which can be analyzed as a p-phrase. This implies that the phrasal contour is clearly realized, especially the final boundary tone of a p-phrase. Every constituent forms its own p-phrase, and all three constituents form an intonation phrase or i-phrase. The final verb and the preceding object are more tightly phrased together than the initial argument. This can be seen in the figures, which show that the high tone of the second argument and the high tone of the verb are fused together. From this high tone, the contour realized on the verb is just smoothly falling until

\(^5\)Although Moore does not mention the effect of downstep it appears to be visible in his data, e.g., his examples (11) or (14) (Moore 1965, 80, 101).
the end of the sentence. The greater prosodic tightness between the last argument and the verb can be expressed as recursive phrasing (see for instance Ito and Mester 2007, for recursive phrasing), as illustrated in (5): The p-phrases of the object and of the verb are grouped together in a single p-phrase that comprises them both. By contrast, and as shown in (6), the order OSV does not have recursive phrasing. A subject and a following verb are separated by a stronger syntactic boundary than an object and a following verb. As a result, they are not grouped together in a common p-phrase. This difference between SOV and OSV order is not reflected in our average data, but we assume that syntax is an important factor in prosodic phrasing, and that it triggers a difference between the two patterns.

(5) a. [[S] [O] [V] I]
   b. [[graahak ne] [[davaai ko] [khariidaa] I] customer ERG medicine ACC buy.PAST

   ‘The customer bought the medicine.’

(6) a. [[O] [S] [V] I]
   b. [[davaai ko] [graahak ne] [khariidaa] I] medicine ACC customer ERG buy.PAST

   ‘The customer bought the medicine.’

The difference between the tonal realization of the p-phrase of the subject and the object on the one hand and verb on the other, comes from the non-finality of the former, and finality of the latter. The tonal structure of the verb is determined by the final low boundary tone (see below), while the tonal structure of the arguments is influenced by the non-final high boundary tone. Importantly for the phonological analysis, phrasing is unchanged by narrow focus, since the phrasal boundaries are always realized, albeit sometimes only weakly. This points to an absence of an effect of focus for phrasing: focus does not insert a prosodic boundary (see similar results for Bengali in Khan 2007, 39–40), though this result needs confirmation from a larger and more varied set of data. This analysis would contradict that of Moore (1965), who assumes a phrase break after a focused constituent.

4.2 Pitch and pitch scaling

Following Nair (1999) and Dyrud (2001) among others, we assume that Hindi has lexical stress, which means that the low part of the rising pattern observed on all non-final constituents can be analyzed as a starred low tone L* for a pitch accent. Harnsberger (1999) proposed that, phonologically, the high part of the rising gesture may be analysed in two different ways: as a high trailing tone (+H-) or as a high phrase tone (HP). Because of the clear phrasing found in our data, we analyze the rising pitch gesture as a low pitch accent L* and a high phrase boundary tone HP, see (7). This is also the pattern proposed by Hayes & Lahiri for Bengali. The final verb has a falling contour, which we assume is coming from a high pitch accent H* and a low boundary tone at the level of the intonation phrase (LI).

(7) L* HP L* HP H* LI

   [[graahak ne] [[davaai ko] [khariidaa] I] customer ERG medicine ACC buy.PAST

Based on the time-normalized pitch contours in Figure 1, we assume that HP is associated with the right edge of the constituent, i.e. the target noun plus case marker. Further evidence for this claim comes from an unpublished study by Genzel (2007) on Hindi. Genzel manipulated the number

---

6 Alternatively, three levels of phrasing may be assumed, as proposed by Khan (2007) for Bengali.
7 The last p-phrase can optionally end with a rising contour which does not necessarily strike Hindi speakers as a list intonation. Some of our speakers regularly realized a rising final intonation, others only occasionally.
8 Moore (1965) reported a similar observation, the rising pitch gesture may continue throughout the noun and any following grammatical morpheme.
A: Wide focus

Figure A shows the phonological interpretation of \([SOV]_F\) (a) and \([OSV]_F\) (b) in the wide focus context; the prosodic phrasing, metrical grid and top lines relationship are displayed. The arrow indicates a lowering of the post-focal pitch range.

B: Subject focus

Figure B shows the phonological interpretation of an SpFOV sentence in subject focus context (initial focus). The prosodic phrasing, metrical grid and top lines relationship are displayed.

C: Object focus

Figure C shows the phonological interpretation of an SOFV sentence in object focus context (preverbal focus). The prosodic phrasing, metrical grid and top lines relationship are displayed.

Her data suggest that the high tone is associated with the right edge of a prosodic word (see Khan 2007, for the same conclusion for Bengali).

The wide focus pattern can serve as a baseline for the remaining contours: All three high tones are subject to downstep. The downstep pattern affects the high tones of prosodic domains. Each high tone is lower than the preceding high tones in the same level of prosodic structure. Figure 4-A shows the metrical structure of sentences of both word orders plus the top lines of the prosodic domains to illustrate pitch scaling. Every constituent is the head of its own prosodic phrase, and for this reason, each constituent has the same metrical level. The downstepped lines above the metrical structure illustrate the top lines of the prosodic phrases. They show the highest point that the speaker’s voice can reach at this moment, and define the maximal height of the high tones. We take the downstep pattern of the p-phrases as an obligatory feature of Hindi intonation. A sequence of prosodic phrases of the same level is organized in downstepped p-phrases.

Narrow focus on one constituent is accompanied by givenness of the other constituents. We represent prominence due to focus and givenness with the help of abstract metrical grid positions. The constituent with narrow focus acquires an additional grid mark, and the following ones become, in relation, less prominent. This difference may change the scaling of tones, as is illustrated with an SpFOV configuration, see Figure 4-B. It is important to realize that the difference in prominence can in principle be realized in two ways implying pitch scaling: either by a rise on the focused constituent, or by a lowering on the given constituents. In Hindi, a change in the focus relationship is expressed by compression of the given constituents, but only of the post-focal ones. The first focused constituent...
does not change its level. When the second constituent is narrowly focused, no difference in scaling appears (Figure 4-C). There is neither raising of the focused constituent nor lowering of the given constituent. We assume that the reason for the total absence of prosodic effects in such a configuration is that the top lines associated with prosodic domains cannot be changed in such a way that downstep inside of an i-phrase is cancelled. This means that the change in the metrical structure as a consequence of the change in information structure has no effect on the relative height of the top lines, and thus, also on the height of the individual tones which are scaled according to these top lines. Raising of the medial constituent would result in suppressing the difference between the first and the second constituent. The same result would appear if the first constituent were compressed. Since both operations would cancel the downstep relation, none of them is performed, and pitch scaling is not affected by focus on the preverbal constituent.

We turn next to word order considerations. First, post-focal compression is larger in SOV than in OSV order. This effect correlates with the difference in phrasing shown in (5) and (6), which correlates with a difference of syntactic boundary strength between a preverbal argument and a verb: it is weak in SOV and strong in OSV sentences. In other words, an object and a following verb are more tightly connected than a subject and a following verb. The difference in post-focal compression reflects this difference. The stronger boundary between a subject and a following verb is also reflected in the significantly longer duration that a focused subject has, as compared to a focused object in the same position (548 ms vs. 528 ms, t=3.56).

Second, in case of preverbal focus, the initial argument displays a greater pitch range and higher F0-maximum in the non-canonical OSV than in the canonical SOV order. This may be the consequence of the scrambled word order in case the object is preposited. As discussed in Section 2, a preposited constituent is generally interpreted as a topic (Gambhir 1981, Butt and King 1996). It may be the case that the speakers, or some of them, have realized the given object as a topic, which would explain the extra high boundary tone. Nevertheless we refrain from analyzing the prosodic phrasing of the initial given object differently from an initial subject, for instance as a separate intonation phrase. There are two reasons for this. First, the context did not give any indication for the speakers to interpret this constituent as a topic; and second, the effects obtained were rather weak. We point to the fact that boundaries of prosodic phrases are subject to gradience anyway, but do not provide a deeper explanation for this effect of word order.

5 Concluding remarks

Based on the Hindi production study, we have proposed that each constituent forms its own prosodic domain, of the size of a prosodic phrase. Non-final p-phrases have a rising pattern (\text{L}^*\text{H}_P) and the final ones have a falling pattern (\text{H}^*\text{L}_I). This structure is not changed by focus. Thus, focus does not introduce a different pattern of phrasing; the prosodic phrases are in a strict downstep relationship which cannot be disturbed.

Hindi differs with respect to the expression of focus and its interaction with the downstep pattern from languages like English or German. In these languages, a sequence of downstepped accents is always interrupted by focus, and focus is realized with an upstep or a raising of the high tone on the focused word (e.g., Bartels and Kingston 1994, Baumann et al. 2006, Cooper et al. 1985, Féry 1993, Féry and Kögler 2008). By contrast, in Hindi focus prominence appears to be expressed after the focused item, by means of post-focal compression. In our material, sentences with focus on the preverbal constituent do not reveal any prosodic difference compared to the wide focus sentences. Since the preverbal position is the syntactic default position for focus (Kidwai 2000), prosodic marking of focus in this position might be redundant and therefore remains unrealized.

Given the analysis above, the global downstep pattern is more important than local register changes induced by focus.

Second, although Hindi’s use of post-focal compression is similar to other languages that reduce the prominence of given material to enhance the salience of focused material (Cruttenden 2006), a
rising pitch accent on content words is required even in post-focal position. This contrasts with other intonational languages, which commonly employ complete deaccentuation of post-focal material.

Third, in contrast to German (and to Bengali, Hayes and Lahiri 1991), pre-focal given elements in Hindi appear to not undergo compression. Given the downstep pattern, it follows that any register compression before a focus would disturb the downstep pattern. If the pitch range of a pre-focal constituent would be compressed, the dissimilative tonal effect of downstep would be blocked, minimizing the prosodic difference between a pre-focal and focal constituent. Downstep seems to be compulsory; we do not find any reduction of the pre-focal pitch register in case of medial focus. It may be that the downstep pattern facilitates sentence comprehension in that it clearly marks constituent boundaries.

It remains to be determined whether the prosodic structure and cues identified here are used by comprehenders to parse sentences more efficiently. Another important open question is whether the constraints identified here are valid for more complex utterances. Answering these questions lies outside the scope of the present study and must be left for future work.

Acknowledgments

We are very grateful to Professor Ramakant Agnihotri for logistical support in running the experiments at Delhi University; without his assistance, none of the work reported here would be possible. Dinah Rottmann and Verena Thiessen assisted with processing the raw data. The data analysis was carried out using the statistical computing language R (R Development Core Team 2006) and the R packages lme4 (Bates and Sarkar 2007). Furthermore, we thank two anonymous reviewers, as well as the editors of JSAL, Miriam Butt and Rajesh Bhatt for helpful comments and suggestions. The research presented here was funded by the Deutsche Forschungsgemeinschaft (DFG), through the project Prosody in Parsing (DFG Schwerpunktprogramm 1234 (2006-2009), FE 292/5-1), Principal Investigators Caroline Féry, Frank Kügler, and Shravan Vasishth.

References

Allerton, David J. 1978. The notion of “givenness” and its relations to presupposition and to theme. *Lingua Amsterdam* 44(2–3):133–168.


Harnsberger, James D. 1999. The role of metrical structure in Hindi intonation. In *South Asian Language Analysis Roundtable, University of Illinois, Urbana-Champaign*.


Keane, Elinor. 2007b. Investigating the Intonational Phonology of Tamil. Poster presented at the satellite meeting of the 16th International Congress of Phonetic Sciences (ICPhS), Saarbruecken, Germany, August 5, 2007.

Khan, Sameer. 2007. Phrasing and Focus in Bengali. Poster presented at the satellite meeting of the 16th International Congress of Phonetic Sciences (ICPhS), Saarbruecken, Germany, August 5, 2007.


6 Appendix

6.1 Tables accompanying the Results section

<table>
<thead>
<tr>
<th>First constituent</th>
<th>Comparison</th>
<th>F0 max (Hz)</th>
<th>F0 range (Hz)</th>
<th>Duration (ms)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Means</td>
<td>t-score</td>
<td>Means</td>
</tr>
<tr>
<td></td>
<td>SOV Wide vs Narrow</td>
<td>274</td>
<td>281</td>
<td>4.26*</td>
</tr>
<tr>
<td></td>
<td>Wide vs Given</td>
<td>274</td>
<td>274</td>
<td>0.15</td>
</tr>
<tr>
<td></td>
<td>Given vs Narrow</td>
<td>274</td>
<td>281</td>
<td>4.15*</td>
</tr>
<tr>
<td></td>
<td>OSV Wide vs Narrow</td>
<td>276</td>
<td>279</td>
<td>1.29</td>
</tr>
<tr>
<td></td>
<td>Wide vs Given</td>
<td>276</td>
<td>279</td>
<td>1.56</td>
</tr>
<tr>
<td></td>
<td>Given vs Narrow</td>
<td>279</td>
<td>279</td>
<td>-0.09</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Second constituent</th>
<th>Comparison</th>
<th>F0 max (Hz)</th>
<th>F0 range (Hz)</th>
<th>Duration (ms)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Means</td>
<td>t-score</td>
<td>Means</td>
</tr>
<tr>
<td>SOV Wide vs Narrow</td>
<td>247</td>
<td>250</td>
<td>1.49</td>
<td>44</td>
</tr>
<tr>
<td>Wide vs Given</td>
<td>247</td>
<td>230</td>
<td>-9.06*</td>
<td>44</td>
</tr>
<tr>
<td>Given vs Narrow</td>
<td>230</td>
<td>250</td>
<td>9.97*</td>
<td>30</td>
</tr>
<tr>
<td>OSV Wide vs Narrow</td>
<td>249</td>
<td>250</td>
<td>&lt;1</td>
<td>42</td>
</tr>
<tr>
<td>Wide vs Given</td>
<td>249</td>
<td>237</td>
<td>-8.23*</td>
<td>42</td>
</tr>
<tr>
<td>Given vs Narrow</td>
<td>237</td>
<td>250</td>
<td>8.65*</td>
<td>32</td>
</tr>
</tbody>
</table>

**TABLE 2** F0-maximum, F0-range, and duration on the first and second constituent for SOV and OSV order in different focus conditions as well as their statistical comparisons by means of t-tests; absolute t-values above 2 are significant and are marked with an asterisk.

<table>
<thead>
<tr>
<th>First constituent</th>
<th>F0 max</th>
<th>F0 range</th>
<th>Duration</th>
</tr>
</thead>
<tbody>
<tr>
<td>Focus</td>
<td>SOV</td>
<td>OSV</td>
<td>t-score</td>
</tr>
<tr>
<td>Wide</td>
<td>274</td>
<td>276</td>
<td>1.5</td>
</tr>
<tr>
<td>Narrow</td>
<td>281</td>
<td>279</td>
<td>&lt;1</td>
</tr>
<tr>
<td>Given</td>
<td>274</td>
<td>279</td>
<td>3.44*</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Second constituent</th>
<th>F0 max</th>
<th>F0 range</th>
<th>Duration</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wide</td>
<td>247</td>
<td>249</td>
<td>1.31</td>
</tr>
<tr>
<td>Narrow</td>
<td>250</td>
<td>250</td>
<td>&lt;1</td>
</tr>
<tr>
<td>Given</td>
<td>230</td>
<td>237</td>
<td>4.55*</td>
</tr>
</tbody>
</table>

**TABLE 3** Maximum F0, duration, and F0-range on the first and second constituent for different focus conditions comparing SOV and OSV word order as well as their statistical comparisons by means of t-tests; absolute t-values above 2 are significant.
6.2  Stimuli
Stressed syllables are capitalized.

6.2.1  Pattern-1 (2 syllabic Subject, 3 syllabic Object)
(1) GRAAhak ne daVAAii ko khariidaa
(The) customer bought the medicine

(2) BAAlak ne suRAAhi ko chhupaayaa
(The) kid hid the jar

(3) GAAyak ne darVAAje ko dhakelaa
(The) singer pushed the door

(4) NAUkar ne kaTOre ko hataayaa
(The) servant took away the bowl

(5) AADmii ne gaVAAhi ko sudhaaraa
(The) man corrected the statement

(6) MAAlik ne kiRAAYe ko badhaayaa
(The) landlord increased the rent

(7) BRAAHi man ne cheTAAVni ko sunaayaa
(The) Brahman announced the warning

(8) SAANsad ne jaanKAari ko failaayaa
(The) parliamentarian spread the awareness

(9) CHHAAttra ne kamPYUter ko ghumaayaa
(The) student turned the computer

6.2.2  Pattern-2 (3syllabic Subject, 2syllabic Object)
(10) maNUshya ne CHAAdar ko jalaayaa
(The) man burnt the bedcover

(11) saVAarii ne JOOte ko utaaraa
(The) passenger took off the shoe

(12) shiKAarii ne PAUdhe ko ukhaadaa
(The) hunter uprooted the plant

(13) khiLAAdii ne GHOde ko bhagaayaa
(The) sportsman made the horse run (faster)

(14) kanDAKtar ne GAAdii ko rukaayaa
(The) conductor stopped the vehicle

(15) shaRAAbii ne BOTal ko bajaayaa
(The) drunkard made sound with the bottle
(16) maiKEnik ne TAxi ko chalaayaa
    (The) mechanic drove the taxi

(17) adHYAapk ne MOORti ko banaayaa
    (The) teacher made the sculpture

(18) adHYAKsha ne PYAAle ko uthaaya
    (The) chairperson picked up the glass
The prosodic expression of contrast in Hindi

Susanne Genzel 1, Frank Kügler2

1 SFB 632 “Information structure”, Potsdam University
2 Department of Linguistics, Potsdam University
susanne7@gmail.com, kuegler@uni-potsdam.de

Abstract
This production study examines the prosodic means of encoding contrast in Hindi. Different target words were embedded in carrier sentences and were put in two information structural contexts, wide and contrastive focus. Contrary to what is expected from earlier findings Hindi uses prosodic means of expressing contrast on the focused word, namely an increase in pitch span and duration. These results may contribute to the understanding of intonational phonology and to the prosodic classification of Hindi.

Index Terms: Hindi, contrast, pitch span, duration

1 Introduction
Hindi is one of the world’s most widely spoken languages with approximately 150 mill. native speakers in north and central India in the states Uttar Pradesh, Uttarakhand, Rajasthan, Bihar, Haryana, Madhya Pradesh, Sikkim, Jharkhand, Chattisgarh und Himachal Pradesh [1]. Hindi belongs to the Indo–Iranian branch of the Indo–European language family, and is genetically related to other European languages like German, English and Russian.

With respect to its prosodic properties most researchers agree that Hindi has lexical stress [2], [3], [4], [5], [6], [7]. The position of stress is driven by syllable weight, whereas the ultima is assumed to be extrametrical [8].

The few studies examining the prosody of Hindi agree that each content word except the last one of an intonation phrase is associated with a rise in pitch [2], [4], [9]. For instance, Moore [2] subdivides sentences into feet. The neutral distribution is one foot per content word, whereas polymorphematic words can be subdivided in two or more feet. According to Moore it is the foot that is prosodically realized with a F0 rise (LH). This indicates that Hindi demarcates the prosodic word prosodically. A study on word order and intonation in Hindi indicated that the constituents in a sentence appear to be in a strict downstep relation [9].

Apart from its demarcative function, intonation also functions as a means of highlighting information that in human communication a speaker transfers to a hearer (e.g. [10]). At the same time other information gets in the background. This function is known as prosodic focus [11], [12], [13]. A focus of a sentence represents a word or constituent that receives prominence either by means of morphology, syntax, prosody or combinations of these linguistic devices. Semantically, focus defines a set of alternatives from which one element is chosen by the speaker [14], [15]. If a focus highlights new information, parts of a sentence may contain already old information which is usually called given information, defined as previously mentioned in the discourse [16], [17]. The information structural notions important for the course of the paper are broad focus and contrastive focus. Following [11] we assume that "... if the focus constituent is the whole sentence, we get 'normal stress'; if not, we get a narrow focus on the constituent identified by the placement of the accent". In the wide or broad focus, all information is new and the whole sentence is focused and serves as baseline condition, also sometimes referred to as out-of-the-blue utterance [19]. In the case of corrective or contrastive focus the speaker has chosen an element from a set of alternatives [14], [15]. The contrastive focus is narrow concerning size [13]. Based on the assumptions of Focus-Prominence-Theory (e.g. [20], [21]) the expression of abstract focus prominence is language specific. Prosodic reflexes due to information structure are well known for intonation languages like German, English and many more. Prominence due to focus is marked by a pitch-accent which shows longer duration and higher F0 whereas given information is deaccented [22], [23], [24]. The greatest amount of focal lengthening in German [25], English [26] and in Swedish [27] can be found in the stressed syllable.

Focus may also be expressed syntactically or morphologically, yet Hindi does not use morphological means to express focus. Hindi is a head-final language with SVO as base word order, and a focused constituent typically would occupy the immediately preverbal position [18]. Since this study is concerned with prosodic reflexes of focus we chose to study the prosody of focus on adjectives. An adjective belongs to a noun phrase, and cannot be split and moved into the preverbal focus position. This strategy ensures that any prosodic effect of focus appears with no other influences.

Previous studies on the effects of focus in Hindi have shown three effects: Greater pitch excursion on the rise, higher intensity and longer duration. In the post–focal region the rise is compressed [2], [4], [28]. In contrast, [9] revealed that a narrow focus in sentence medial position showed no difference in high tone scaling as compared to the all-new pattern. Given these contradictory findings this study will examine the prosodic expression of contrast in Hindi in more detail.

2 Method

2.1 Speech materials
Table 1 shows the different adjectives used as target words. The adjectives differed in number of syllables as well as in position of stress. These were put in carrier sentences with SVO structures modifying the object (1).

Wide focus was elicited without a context, i.e. out of the blue [19]. The contrastive focus elicitation consisted of a question-answer pair: a question and a response to the question (1). The focus domain is indicated by squared brackets and the F mark.
Trials for the contrastive focus consisted of written and verbal instructions, followed by four practice trials. The microphone was an electret condenser type with a sensitivity of -39 dBV/Pascal.

The experiment was carried out using presentation software. The whole presentation was carried out in a quiet room at their houses or working places using a headset (Logitech Internet Chat Headset). The headphones were binaural with a frequency spectrum from 20-20000 Hz and an acoustic impedance of 32 Ohm with an integrated volume control, so that every participant could adjust the volume. The acoustic impedance of 32 Ohm with an integrated volume control, so that every participant could adjust the volume.

The participants were familiarized with the task through written and verbal instructions, followed by four practice trials. Trials for the contrastive focus consisted of a visual presentation of the question and its answer on the computer screen. For elicitation of the wide focus condition only the presentation of the question and answer on the computer screen was present, and if not, at the beginning of the actual pitch rise. To normalize for pitch range and sex, two reference points were set, one for high pitch (R1) before the target word and one for low pitch (R2) after it. Obvious errors in the pitch track were manually corrected, and F0 was smoothed at 10Hz. The pitch values were extracted via a Praat script. The actual Hz (x) values were normalized applying the formula in (3) following [31].

\[ y = \frac{x - R2}{R1 - R2} \]

The statistical analysis relied on the dependent variables 'word duration' and 'F0' calculated for L, H and span. A paired samples T-test was carried out in R. A p-value smaller than 0.05 indicates statistical significance.

### 2.3 Acoustic analysis

The recordings were digitized at a sampling frequency of 44.1 kHz and 32 bit resolution. The target word in all 814 sentences (37 target words \( \times 2 \) information structural conditions \( \times 11 \) speakers) was labeled by hand at the level of the word, the syllable, and the segments. Labeling in Praat [29] was based on a see-listen-label method, visually evaluating the spectrogram of the sound files and listening. Standard cues for segmental labeling were employed [30], and boundaries were set automatically at zero crossings via a Praat script. For each target word, the duration of the word, the syllables and the segments was extracted in ms using a Praat script.

In the target word the rise was labelled. The H tone was set at the local pitch maximum, the L tone at the local low if present, and if not, at the beginning of the actual pitch rise. To normalize for pitch range and sex, two reference points were set, one for high pitch (R1) before the target word and one for low pitch (R2) after it. Obvious errors in the pitch track were manually corrected, and F0 was smoothed at 10Hz. The pitch values were extracted via a Praat script. The actual Hz (x) values were normalized applying the formula in (3) following [31].

\[ y = \frac{x - R2}{R1 - R2} \]

### 2.4 Statistical analysis

The statistical analysis relied on the dependent variables 'word duration' and 'F0' calculated for L, H and span. A paired samples T-test was carried out in R. A p-value smaller than 0.05 indicates statistical significance.

### 3 Results

#### 3.1 Information structural effects on syllable duration

The results from the statistical analysis reveal that contrastive focus significantly affects the duration of the target words in Hindi, \( t(10)=2.2281, p = 0.01 \). To gain a better understanding of focal lengthening in the words Table 2 presents the mean syllable duration for both information structural conditions, with wide focus (wFoc) as baseline condition compared to contrastive focus (cFoc). From Table 2 it can be seen that in contrastive focus each syllable of a target word is lengthened compared to wide focus renditions.
difference. The smallest contribution is made by the second syllable which contributes the biggest amount (35%) of the durational difference to the two and three syllable target word. The stressed item only one, the whole word gets lengthened by 23%. For the initially stressed four syllable target word, which contains unstressed items all syllables contribute nearly equally (24-27%) to the lengthening of the words which amounts 51%. In the unstressed syllable 2%, the third syllable is lengthened by 33% and the final one 20%. For the two remaining 4 syllable target words all unstressed syllables are not involved in the durational difference.

### Table 3: Number of syllables (nos), stress pattern (0=unstressed, 1=stressed), amount of durational change in ms ($\Delta C$), and as percentage (%L) for word and syllable level, the syllable-to-word-lengthening-ratio (%$\Delta C$/%$\Delta W$) for contrastive focus compared to wide focus.

<table>
<thead>
<tr>
<th>nos</th>
<th>stress</th>
<th>$\Delta C W$</th>
<th>$\Delta C r$</th>
<th>%$\Delta C$/%$\Delta W$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>15</td>
<td>15</td>
<td>100</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>11</td>
<td>3</td>
<td>5</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>11</td>
<td>3</td>
<td>10</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>43</td>
<td>10</td>
<td>51</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>43</td>
<td>10</td>
<td>14</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>43</td>
<td>10</td>
<td>35</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>37</td>
<td>8</td>
<td>32</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>37</td>
<td>8</td>
<td>43</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>37</td>
<td>8</td>
<td>24</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>110</td>
<td>23</td>
<td>35</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>110</td>
<td>23</td>
<td>25</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>110</td>
<td>23</td>
<td>33</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>110</td>
<td>23</td>
<td>30</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>110</td>
<td>23</td>
<td>40</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>51</td>
<td>10</td>
<td>22</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>51</td>
<td>10</td>
<td>25</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>51</td>
<td>10</td>
<td>27</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>13</td>
<td>2</td>
<td>62</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>13</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>13</td>
<td>2</td>
<td>6</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>13</td>
<td>2</td>
<td>0</td>
</tr>
</tbody>
</table>

For target words with four syllables all syllables except one appear to contribute equally to word lengthening. For the initially stressed four syllable target word, which contains only one item, the whole word gets lengthened by 23%. Equally to the two and three syllable target word the stressed syllable contributes the biggest amount 35% of the durational difference. The smallest contribution is made by the second unstressed syllable 2%, the third syllable is lengthened by 33% and the final one 20%. For the two remaining 4 syllable target words all unstressed items all syllables contribute nearly equally (24-27%) to the lengthening of the words which amounts 51%. In the penultima stressed four syllable target words which are lengthened by 13% the stressed syllable contributes only 15% to the total amount of lengthening. The first syllable contributes 62%, the second one 23% and the final syllable is not involved in the durational difference.

### 3.2 Information structural effects on F0

Figure 1 shows the mean normalized $F_0$ values for the L, the H tone and pitch span, which is calculated as the difference between the two tones [13]. The L tone gets significantly lower under contrastive focus ($t(9) = 2.301, p < .05$), whereas the H tone gets significantly higher ($t(9) = 4.130, p < .01$). As a consequence also the pitch span is significantly enhanced under contrastive focus ($t(9) = 4.613, p = .001$).

To get an insight into speaker specific strategies Table 4 shows the mean pitch change for L and H tones between contrastive and wide focus for all speakers. Concerning the L tone Table 4 reveals that only half of the speakers lower the L tone. This is true for both females and males. All speakers, however, increase the height of the H tone.

### Table 4: Mean amount of pitch change in normHz ($\Delta C$) for L and H for contrastive focus compared to wide focus, for all subjects, sex.

<table>
<thead>
<tr>
<th>subject</th>
<th>sex</th>
<th>$\Delta C L$</th>
<th>$\Delta C H$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>♂</td>
<td>-0.32</td>
<td>0.27</td>
</tr>
<tr>
<td>2</td>
<td>♂</td>
<td>-0.30</td>
<td>1.16</td>
</tr>
<tr>
<td>3</td>
<td>♂</td>
<td>0.06</td>
<td>0.88</td>
</tr>
<tr>
<td>4</td>
<td>♂</td>
<td>-0.45</td>
<td>0.08</td>
</tr>
<tr>
<td>5</td>
<td>♂</td>
<td>-0.04</td>
<td>0.32</td>
</tr>
<tr>
<td>6</td>
<td>♂</td>
<td>0.00</td>
<td>0.42</td>
</tr>
<tr>
<td>7</td>
<td>♂</td>
<td>-0.14</td>
<td>0.56</td>
</tr>
<tr>
<td>8</td>
<td>♂</td>
<td>0.05</td>
<td>0.17</td>
</tr>
<tr>
<td>9</td>
<td>♂</td>
<td>-0.33</td>
<td>0.56</td>
</tr>
<tr>
<td>10</td>
<td>♂</td>
<td>-0.32</td>
<td>0.27</td>
</tr>
</tbody>
</table>

Any change in tonal realization results in a larger pitch span, and speakers differ as to which tone mostly contributes in the change of pitch span. Most speakers lower L tones and raise H tones at the same time. Speaker 4 only lowers the low tone with almost no increase of the H tone, and speakers 3, 7 and 9 only raise the H tone to achieve a larger pitch span.

### 4 Discussion

The results of our experiment suggest that contrastive focus affects the pitch rise associated with a prosodic word in Hindi declaratives in two ways. First, duration is affected. The word is lengthened under contrastive focus. The involvement of each syllable was studied in detail because stress in Hindi is not an uncontroversial issue. Most researchers agree that there
is stress ([2], [3], [4], [5], [6], [7]), the phonetic cues and listeners agreement on the placement of stress are yet still unclear. Ohala [3] suggests that stress is far weaker than in English, but for many words there is only one syllable that can be made prominent. Regarding the involvement of the stressed syllable in the focal lengthening the results show that for diastratic syllable target words the stressed syllable contributes the biggest amount to the word lengthening effect. The same effect was reported for the four syllable target word with initial stress, though not for the other four syllable words. Putting these words aside there seems to be a hint that quantitatively strong syllables are more involved in the lengthening than the unstressed syllables like in German [25] and Swedish [27], a result which may contribute to the understanding of Hindi stress. In non-stress-accents languages like Japanese only F2 should be used as a phonetic correlate of accentual prominence [32]. It would be worth studying the lengthening in more detail looking for segmental and syllable structure effects as well as for domain effects.

Second, the scaling of the pitch contour is affected. Contrary to [9] a higher scaling for the H tone in contrastive focus was observed in comparison to the wide focus baseline. And the L tone was also affected showing a significant lowering due to contrastive focus which together with a change of the H tone results in an increased pitch span. Concerning speaker variation we showed that all speakers modulated pitch span but to a different extend to express contrast prosodically. The tonal configuration is not changed under contrastive focus. Speaker specific strategies expressing focus are also reported for German [33]. A perception experiment testing the relevance of the phonetic cues found in this study would shed light on the perceptual utilization and maybe also on functional ranking of the cues.

Apart from the well established H raising effect in contrastive focus reported for intonation languages [17], [22], [26], [24] this study shows a reverse effect for the L tones which results in a difference in pitch span. The tonal distinctions are made sharper, an effect also reported for Mandarin Chinese [34]. Thus, it is the strategy of pitch span change (lowering L tones, raising H tones, or both) that makes Hindi an interesting case in terms of prosodic typology regarding the expression of focus.

5 Acknowledgements

Thanks to Anjana Singh for preparation of the target words, to Shravan Vasishth for taping of the questions, Umesh Patil for preparation of the material in Devenagri script and all participants attending. This research has been funded by a DFG research grant (SFB 632/D5) at Potsdam University.

6 References

Section III

Yucatec Maya
Interaction of Lexical Tone and Information Structure in Yucatec Maya

Frank Kügler & Stavros Skopeteas
Department of Linguistics
University of Potsdam, Germany
{kuegler;skopetea}@uni-potsdam.de

Abstract

Previous research on the tone system of Yucatec Maya provides contradictory accounts which this paper intends to do away with, disentangling tonal and intonational effects. The first part presents the mere realisation of lexical high and low tones, the only tonal distinction we identify for Yucatec Maya. Second, we claim that in Yucatec Maya no interaction exists between intonation and lexical tone. We prove this claim showing that neither topic nor focus is realized by means of intonational pitch accents; instead they are marked only by syntax. Deviating tonal patterns from tonal default realisation are a result of tonal effects that surface as tonal transitions, and/or phrasing effects.

1. Preliminaries

1.1. Lexical tones in Yucatec Maya

Yucatec Maya is a Mayan language spoken in Yucatecan Peninsula (Yucatán, Quintana Roo, Campeche, and also in Belize). Among the contemporary Mayan languages, Yucatec Maya is spoken by the largest population (700,000 speakers according to the 1990 census). Yucatec Maya is the only Mayan language that displays lexical tones. According to the reconstruction in [6], tonogenesis took place already in Proto-Yucatecan, which contains three other Mayan languages, namely Mopan, Itzà, and Lacandon.

It has been argued on the basis of a sparse data base in [1] that the distinction of tones is extinct in currently spoken Yucatec Maya. A complete loss of tone is not reported by other investigations and this is in line with our experience in the field: apart from the uncertainty of certain speakers concerning the tone of particular lexical items, the tonal distinction is active in the language production of older and younger speakers.

There are several and partly controversial accounts about the tonal system of modern Yucatec Maya. The phoneme inventory displays a distinction between short and long vowels. All investigations agree that long vowels are obligatory tone bearing units and display an opposition between a high tone and a low tone. Short vowels are treated as contrasting two levels of pitch in [9], or as instantiating a third tone termed as “neutral” in [6], or as having no tone in [2]. The tonal distinction as well as the distinction between long and short vowels is shown to be contrastive: lük’ul ‘goes away’ - lük’ul’ ‘swallow’ - lük’ul’ ‘mud’ (examples from [7]; see also [2] and [9]).

Concerning the phonetic realisations, the lexical low tone is a level tone according to [2], [9], and [10]. The lexical high tone is described as a rising tone in [2], but as a falling tone in [6]. However, [6] shows that the falling realisation occurs in monosyllabic words while in the first syllable of disyllabic words the lexical tone is realized as a rise, and [10] treats the rising contour of the high tone as its indispensable part in the different phonetic realisations. [9] identifies two realisations of lexical high tone, either “falling from a high pitch” or “remaining at a high pitch”. None of these investigations argues that the several realisations of high tone are contrastive at the lexical level.

Properties of Yucatec Maya intonation are dealt with in [2], which offers a detailed annotation of intonational contours made for didactic purposes. Furthermore, [10] gives an inventory of rules that predict different realisations of the lexical tones in several tonal environments.

1.2. Some remarks on Yucatec Mayan syntax

Since a part of this paper is devoted to the prosodic correlates of information structure, some remarks on the syntax are necessary. According to [5] and [11] Yucatec Maya is a head marking VOS language as can be seen in (1).

\[ t-u \ h\=aant-ah \ \hat{o}n \ \text{Pedro.} \]  
(1) PFV-A.3 eat:TRR-CMPL(B.3.SG) avocado Pedro 'Pedro ate avocado.'

Topicalisation and focusing are indicated by movement to designated topic and focus positions, respectively. The topic constituent is left dislocated (see [3]), its right boundary is marked by the suffix -e’, as illustrated in (2). Arguments as well as non-arguments may be topocalized. Thus, the topic position may be occupied by nouns, pronouns, adverbs, adjectives, and clauses (see [3]). Multiplex topics are also usual in spontaneous discourse (cf. test sentences A1c and A2c in the Appendix).

\[ \text{Pedro-}\hat{e}’ \ t-u \ h\=aant-ah \ \hat{o}n \ ]  
(2) Pedro-TOP PFV-A.3 eat:TRR-CMPL(B.3.SG) avocado 'As for Pedro, he ate avocado.'

Focus assignment is expressed by the displacement of an argument in the preverbal domain (compare (3a) with (1)). Such argument focus constructions with preverbal focus are analyzed as cleft constructions (see [4]).

\[ \hat{o}n \ t-u \ h\=aant-ah \ \text{Pedro.} \]  
(3a) avocado PFV-A.3 eat:TRR-CMPL(B.3.SG) Pedro 'It was an avocado, that Pedro ate.'

\[ \text{Pedro} \ h\=aant \ \hat{o}n. \]  
(3b) Pedro eat:TRR(SUBJ)(B.3.SG) avocado 'It was Pedro, that ate an avocado.'

In (3b), agent focus is illustrated, which is expressed through a special ‘out of focus’ form of the verb. The aspect
auxiliary is dropped together with the cross-reference clitic for the agent. In the perfective aspect (3b), the extrafocal verb bears the zero form subjunctive marker in non-clause-final position.

2. Speech materials

2.1. Resources

The data presented in this paper was collected during our field work in December 2004 in Quintana Roo, Mexico. Our informants live in a community of about 800 speakers (Yaxley, Quintana Roo), and mainly use Yucatec Maya in their everyday communication, although all are bilingual in Spanish.

In total, twelve speakers have been recorded. However, all twelve speakers did not produce sentences with all test items except for the minimal pair míis ‘broom’ and mìiis ‘cat’.

2.2. Description of the production experiment

Speech Materials. Since information structure is encoded through particular syntactic structures in Yucatec Maya, the first question is if the topicalisation and focusing constructions illustrated in section 1.2 are associated with particular tonal events. In order to isolate tonal events associated with information structure and lexical tones, we have developed a small text containing the three constructions under investigation (cf. Appendix). Two versions of this text have been used, one for animate (cf. A1) and one for inanimate target words (cf. A2): (a) a sentence with the target word as a single argument of the existential verb (broad focus condition; see A1a and A2a); (b) a sentence with the target word in the focus position (narrow focus condition; see A1b and A2b); (c) a sentence with the target word in the topic position (topic condition; see A1c and A2c).

It should be noticed that Yucatec Maya is one of the languages that encode discourse functions through syntactic constructions and morphological marking (see section 1.2), hence comparing the tonal realisation sentences which are morpho-syntactically identical but differ in information structure is not allowed for by the structure of the language. By consequence, the target elements are necessarily embedded in different sentences specifically chosen to allow for observation of (possible) tonal events that accompany the morpho-syntactic structures that are related to information structure. In all sentences the target words are non-initial and morpho-syntactic structures that are related to information embedded in different sentences specifically chosen to allow for observation of all possible tonal patterns (see Table 1). In this article, we discuss just some representative cases of the tonal phenomena at issue.

Data elicitation. The speech data were elicited by means of question-answer pairs. Since most Yucatec Mayan speakers are not trained in reading Mayan orthography, we had to present our stimuli orally. The carrier sentences with target items as given in Table 1 were thus read by a native speaker before running the experimental sessions. The pitch contour of each provided sentence, however, has been reduced to a flat level pitch in order to eliminate all linguistic information that is encoded by pitch. In the experimental sessions, informants heard the resynthesized stimuli. The informants’ task, then, was to answer a generic question by repeating the text they had just heard before. All recordings were made on a DAT recorder (SONY 100) using head microphones. For the manipulation of the test sentences and for pitch analyses we used Praat (see [13]).

Table 1: Tonal patterns in lexical items.

<table>
<thead>
<tr>
<th>tonal pattern</th>
<th>lexical item</th>
<th>translation</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>am</td>
<td>spider</td>
</tr>
<tr>
<td>L</td>
<td>lóol</td>
<td>flower</td>
</tr>
<tr>
<td>L</td>
<td>mìis</td>
<td>cat</td>
</tr>
<tr>
<td>H</td>
<td>mìis</td>
<td>broom</td>
</tr>
<tr>
<td>H</td>
<td>láal</td>
<td>stinging nettle</td>
</tr>
<tr>
<td>N-N</td>
<td>ahaw</td>
<td>chief</td>
</tr>
<tr>
<td>N-L</td>
<td>konkluum</td>
<td>pot seller</td>
</tr>
<tr>
<td>N-H</td>
<td>konchuíuk</td>
<td>shoe seller</td>
</tr>
<tr>
<td>L-N</td>
<td>yiíyuum</td>
<td>bird</td>
</tr>
<tr>
<td>L-H</td>
<td>kóolnáal</td>
<td>farmer</td>
</tr>
<tr>
<td>L-L</td>
<td>xííuxkuúts</td>
<td>pheasant</td>
</tr>
<tr>
<td>H-N</td>
<td>yáalam</td>
<td>fawn</td>
</tr>
<tr>
<td>H-L</td>
<td>óochkàaan</td>
<td>snake</td>
</tr>
<tr>
<td>H-H</td>
<td>töochchúuk</td>
<td>coal merchant</td>
</tr>
</tbody>
</table>

3. The realisation of lexical tones

The first observation to be made is that older speakers as well as younger ones exhibit tone in their grammar of Yucatec Maya – in contrast to the observations in [1]. On syllables containing a long vowel we identify a tonal distinction between an underlying high (H) and underlying low tone (L), which is in line with [2], [6], and [9]. In addition, syllables containing a short vowel are toneless underlyingly (= neutral, cf. Table 1). In the following, we provide a qualitative overview of the data comparing monosyllabic with disyllabic target words that bear a low or a high tone, or a combination of the two according to Table 1.

3.1. The lexical low tone

A lexical low tone in Yucatec Maya is realized with low level pitch. As can be seen in Figure 1, the monosyllabic target word lóol ‘flower’ is pronounced with flat pitch at a constant level. The rise in pitch at the end of the target word is due to a high tone associated with the topic marker –é’.

Similarly, a disyllabic target word with a low tone on the first and a high tone on the second syllable, i.e. kóolnáal ‘farmer’, is realized with low pitch on the first and a rising pitch on the second syllable (cf. Fig. 2). Comparing the low tone of Fig. 1 with that of Fig. 2, we observe that the former is low flat while the latter is realized slightly falling. The additional pitch height at the onset of the first syllable’s vowel is due to the segmental context of the syllable’s onset, i.e. the unvoiced velar plosive [k] raises the pitch. Thus, the slightly falling realisation can be explained as a microprosodic effect. Further, in case of a following topic marker as in Fig. 2, an additional rise due to the high tone associated with the topic marker can be observed. To conclude, our data verifies the view in [2] and [9] that the lexical L is a level tone.
Figure 1: Target word lòol 'flower' with lexical low tone in topic position; cf. sentence frame (A2c).

Figure 2: Target word kòolnáal 'farmer' with lexical low on the first and high tone on the second syllable in topic position; cf. sentence frame (A1c).

Figure 3: Target word láal 'stinging nettle' with lexical high tone in topic position; cf. sentence frame (A2c).

Figure 4: Target word konchúuk 'shoe seller' with lexical high tone on the second syllable in broad focus; cf. sentence frame (A1a).

3.2. The lexical high tone

A lexical high tone in Yucatec Maya is realized with a rise in pitch approaching a high tonal target, cf. Fig. 3. The rise starts from a low pitch level that is equivalent to the low pitch levels of the sentence initial word ku ts’o’kol-e’ ‘afterwards’, about 140 Hz for the particular speaker in Fig. 3. From that level at the onset of the target word, the pitch rises about 23 Hz. A preliminary analysis of the rise for four speakers reveals a mean rise of 1.99 semitones.

A comparison of the target word of Fig. 3 with a disyllabic target word containing a lexical high tone on the second syllable while the first one is tonally unspecified (konchúuk ‘shoe seller’) reveals, again, that a high tone is realized similarly as in a monosyllabic word, cf. Fig. 4. In order to implement a rise, the pitch on a preceding syllable falls to a low target level. In Fig. 4, the rise is interrupted due to the unvoiced segmental context of the second syllable’s onset, yet the target of the high tone is clearly visible on the nucleus of the second syllable. The pitch falls gradually after the target word towards the end of the intonation phrase.

The present data suggest that a lexical high tone is realised as rising, which is in line with [2] and [10]. However, considering the disyllabic word tóokchúuk ‘coal merchant’ with each syllable associated with a lexical high tone, we observe a rise in pitch for only one of four speakers. Three speakers realise these two successive high tones as high level pitch (cf. Fig. 11 below). What we may conclude, however, is that the view of [6] and [9], who claim falling pitch for high tones, is refuted. A Yucatec Mayan high tone appears to be realised as rising or high level pitch but not falling (for a discussion of falling pitch in combination with a high lexical tone, see below section 4.3).

4. Yucatec Mayan Intonation

4.1. Focus in Yucatec Maya

As illustrated in section 1.2, narrow focused constituents appear preverbally (cf. sentence frames (A1b) and (A2b)). If a word containing a lexical prespecified tone occurs in the focus position, the underlying shape of the tone as described in sections 3.1 and 3.2 remains preserved. Any deviation from the underlying pattern may be explained by tonal effects, such as tonal transitions, and/or phrasing. Thus, we observe no interaction of lexical tone and intonation, in particular pitch accents for the expression of focus (see section 4.2 below for a
In Fig. 5, a pitch track of the monosyllabic target word *míis* ‘broom’ in narrow focus position is shown. The target word is realized with the rise in pitch that characterizes a lexical high tone (see section 3.2). There appears no further tonal event that might be analysed as a pitch accent indicating focus tonally. If we compare the narrow focus realisation of a target word containing a lexical high tone with a realisation in broad focus (postverbally) or in topic position (preverbally as in the narrow focus condition, cf. Fig. 6), we observe the same tonal pattern, i.e., a rise in pitch on the target word (cf. Figs. 5, 6, and 7). Thus, we may conclude that information structural components such as topic, narrow and broad focus are not expressed by means of post-lexical tones (pitch accents) as is the case in intonation languages such as English (cf. [12]).

If we compare different instantiations of the low tone realised on the same target word (here: *lòol* ‘flower’), we observe that the realisation in Fig. 8 (narrow focus) corresponds to the properties of lexical low tones as illustrated in section 3.1, but the realisation in Fig. 9 (broad focus) displays an unexpected fall in pitch. Yet, we argue that broad and narrow focus are not distinguished tonally. The difference in the observed contours is due to a difference in discussion of a possible phrase tone as a result of the topic marker).

**Figure 5:** Target word *míis* ‘broom’ with lexical high tone in narrow focus; cf. sentence frame (A2b).

**Figure 6:** Target word *míis* ‘broom’ with lexical high tone in topic position; cf. sentence frame (A2c).

**Figure 7:** Target word *míis* ‘broom’ with lexical high tone in broad focus; cf. sentence frame (A2a).

**Figure 8:** Target word *lòol* ‘flower’ with lexical low tone in narrow focus; cf. sentence frame (A2b).

**Figure 9:** Target word *lòol* ‘flower’ with lexical low tone in broad focus; cf. sentence frame (A2a).
phrasing. In case of Fig. 8 (narrow focus), a clear phrase break prior to the target item occurs, whereas in case of Fig. 9 (broad focus), the phrase break occurs first after the target word. To reach the low target of the word bōol ‘flower’ in Fig. 9, a tonal transition arises between the previous lexical high tone on the inanimate indefinite hunpēel and the following target word. Thus, the greater fall in pitch is not a characteristic of the low tone itself, nor is it directly due to a difference of information structure. If phrased differently, i.e. with a pause prior to the target word, we would expect the low tone in Fig. 9 to be similar to that of Fig. 8.

A similar effect arises when a disyllabic target word with a high tone associated with the second syllable follows the animate indefinite huntial, cf. Fig. 4. The tonal sequence of two H-tones is interrupted by a syllable with no lexical tone. The pitch on that syllable is a mere transition, and the fall resembles the fall towards a low tonal target as in Fig. 9. This strengthens our basic assumption concerning the realisation of lexical L-tones. Whenever a falling realisation occurs in our corpus, it may be accounted for through a preceding high target.

In sum, our data does not provide evidence for tonal events associated with the focus position. Comparing the realisation of narrow focus with that of a topic, no differences can be observed (cf. Figs. 1 and 8). As for the lexical high tone associated with words in different information structural positions, in case of a target word containing a lexical low tone we observe no tonal event that might be analysed as a post-lexical tone (pitch accent) to express topic or focus.

4.2. Topic in Yucatec Maya

As [1], [3], and [7] observed, topics are left dislocated in Yucatec Maya, and the topicalized constituent is obligatorily marked with a topic suffix. As can be observed in Figs. 1, 2, 3, 5, 6, and 7, topicalized constituents – either single or multiplex topics – are accompanied by a salient tonal event: a high tone associated with the right edge of the topic phrase.

There are four possible hypotheses about the status of this tonal event: (i) it is related to information structure, thus marking a phrase as topic (in this case it would be a boundary tone); (ii) it is a lexical high tone associated with the suffix -e’; (iii) it is the result of the phrase boundary, thus being a boundary tone, but in contrast to (i) it is independent of the information structure; and (iv) it is associated with the glottal stop. According to hypothesis (ii) this high target is a lexical tone, according to hypotheses (i) and (iii) it is a postlexical tone, and according to hypothesis (iv) it is conditioned by a phonological segment. We have stated in section 3 that lexical tones are associated with long vowels, so the occurrence of a lexical tone on the suffix -e’ would violate the general principles of tonal association in Yucatec Maya.

In case of two successive topicalized constituents, the effect of tonal upstep can be accounted for as a result of a sequence of two high tones. Consider Fig. 2, for instance. The tonal sequence of L-H-H causes an upstep of the second H tone, which is associated with the topic suffix. The same effect is shown in Fig. 3, where the second high tone of a H-H sequence is realized higher than the first. Again the second high tone is associated with the topic suffix -e’, and in both cases (see Figs. 2 and 3), the second topic suffix is realized at the same pitch level as the first one.

In case of a low tone preceding the topic suffix (i.e., a tonal sequence of L-H), we observe a similar rise due to the high tone associated with the suffix (cf. Fig. 1). The crucial difference between a L-H and a H-H sequence is that in the former case, the pitch level of the second topic suffix is lower than the first (cf. Fig. 1).

However, if two lexical high tones occur in the same word (see Fig. 11, right panel), no upstep of the second high tone occurs. The pattern shown in Fig. 11 has been produced by three of four speakers. As mentioned in section 3.2, the fourth speaker realised the high tones as rising ones, yet also with no upstep of the second high tone (cf. Fig. 10).

4.2.1. Topic in Yucatec Maya

Based on the observation that sequences of high tones show different tonal behavior, i.e., upstep or no upstep, we might assume that in case of upstep, two different types of tones are involved. Given that lexical tones are not upstepped as Figs. 10 and 11 show, we draw the conclusion that the second high tone in Figs. 1, 2, 3, and 6 is not a lexical one, what is in line with the phonological restriction that tone bearing units be long vowels in Yucatec Maya. This piece of evidence supports the exclusion of hypothesis (ii), according to which the high tone at the right edge of topic constituents is a lexical tone.

The suffix -e’ belongs to a class of suffixes that display the same phonological structure, the local deictic suffixes -a’ ‘D1’, -o’ ‘D2’, and the negative enclitic -i’ ‘NEGFI’. All these elements occur phrase finally and are associated with the same tonal events as the topic suffix -e’. The tonal behavior of these elements may be observed in Fig. 7 (see the high tone associated with the right boundary of the final phrase). The realisation of the high tone in the environment in which we would expect a final lowering is not obligatory, but it is a characteristic property of IPs ending with suffixes of this class. On the basis of this evidence we can rule out hypothesis (i), that this tonal event is associated with the discourse function of topic phrases.

In sum, we have given empirical evidence that the high tone occurring at the right edge of topic phrases is neither a lexical tone nor a boundary tone related to the information structure of these constituents. Our experimental study does not provide conclusive evidence to decide between hypothesis (iii) that the high tone is a postlexical tone associated with a type of IP or (iv) that the high tone is associated with the glottal stop, since hypothesis (iv) requires the examination of items with a final glottal stop that were not part of our sample (see Table 1).
In line with the conclusions in section 4.1, our analysis of topic constituents shows that the corresponding tonal events are not triggered by the information structure, but relate either to phrasing or to phonological conditions.

### 4.3. Boundary tones in Yucatec Maya

In Yucatec Maya, we observe tonal phenomena that we might analyse as boundary tones. We have already argued in section 4.2 for a possible high boundary tone that delimits phrases ending to a special class of enclitics. In this section, we discuss the instance of a low phrase boundary tone that interacts with a lexical high tone.

According to [9], it remains unclear whether a distinction between a falling and a high level lexical tones exists. Our analysis provides evidence against such a claim, i.e. we assume that a high tone may fall if the tone-bearing unit happens to occur phrase finally. The left panel of Fig. 11 displays a one-word phrase taken from a spontaneous discussion with one of the informants, who explains the target word several times in isolation. Given that a so-called citation form forms its own intonation phrase (e.g. [12]), we analyse the tonal fall in this particular case as an interaction between a lexical high tone and a low intonation phrase boundary, thus an interaction between tone and intonation. If the target word is not phrase-final (cf. right panel of Fig. 11), both high tones are realized high, thus no fall is produced.

![Figure 11: Target word tóokchúuk 'coal merchant' with two lexical high tones in a one-word phrase (left panel) and extracted from frame (A1a) (right panel).](image)

### 5. Acknowledgments

We wish to thank Caroline Féry, Christian Lehmann, and Elisabeth Verhoeven for their advices, discussion, and comments during the preparation of the field work and the analysis of Mayan tones and intonation, and Caroline Magister and Elizabeth Medvedovsky for their assistance in the preparation of the article. This paper is part of the project D2 of the SFB 632 on Information Structure at the University of Potsdam, financed by the DFG.

### 6. References


ON THE UNIVERSALITY OF PROSODIC REFLEXES OF CONTRAST: THE CASE OF YUCATEC MAYA

Frank Kügler & Stavros Skopeteas

Department of Linguistics, Potsdam University
f.kuegler@googlemail.com; skopetea@uni-potsdam.de

ABSTRACT

This paper is about the tonal realization of contrastive focus in Yucatec Maya. Examining sentences with in situ focused adjectives (postverbally) we observe neither durational differences as compared to non-contrastive sentences nor any differences in F0 except for a higher non-contrastive high tone. Yucatec Maya, being a tone language, seems to use prosodic means exclusively to express tonal contrasts, thus belonging to a language type without prosodic marking of contrast.

Keywords: Yucatec Maya, contrastive focus, prosody, tone.

1. INTRODUCTION

Yucatec Maya is a Mayan language spoken by a population of 700,000 speakers (following the 1990 census) in the Yucatecan peninsula. It is the only Mayan language that displays lexical tones. Yet, as for many Mayan languages, the basic word order is V-initial with postverbal OS order, e.g. [8].

Yucatec Maya displays several morphosyntactic means of encoding information structure including topic and focus constructions, topic affixes, and out-of-focus verb morphology [5]. Regarding contrastive focus, [3] does not observe any prosodic marking in sentence final position. Similarly, [4] observe no prosodic marking in the comparison of broad and narrow focus. The present study examines contrastive focus in postverbal but sentence medial position, in contrast to [3], excluding possible influences of finality which may have “overwritten” the prosodic marking of contrastive focus in [3].

1.1. Syntactic constructions

As is the case for most languages of the Mayan family, the basic word order is V-initial. The order of postverbal arguments is basically OS (see (1) and [8]).

(1) t-u hàant-ah òon Pedro.  
PFV-A.3 eat:TRR-CMPL(B.3.SG) avocado Pedro  
‘Pedro ate avocado.’

Focused constituents are placed left adjacent to the verb (compare (2a-b) with (1)). A difference in the morphology creates an agent/patient asymmetry: only in agent-focus (and in perfective aspect) is the aspect auxiliary dropped, together with the cross-reference clitic for the agent. This verb form is non-finite (also characterized as an ‘out of focus’ verb form) and never occurs in main clauses. Accordingly, such argument focus constructions are analyzed as cleft sentences [9].

(2) a. òon t-u hàant-ah Pedro.  
avocado PFV-A.3 eat:TRR-CMPL(B.3.SG) Pedro  
‘It was an avocado, that Pedro ate.’

b. Pedro hàant òon.  
Pedro eat:TRR(SUBJ)(B.3.SG) avocado  
‘It was Pedro, that ate an avocado.’

Data from production experiments revealed an asymmetry in the obligatoriness of the focus construction as in (2). While there is a strong preference to express focus on the agent constituent through the focus construction, focus on the patient was encoded either through the focus construction or in situ, with almost the same frequency [5]. The crucial generalization for our purposes is that in situ placement of the patient constituent is not specified syntactically for information structure. Consequently, this syntactic configuration is suitable for inspecting the availability of prosodic reflexes of information structure, which is the aim of the present study.

1.2. Yucatec Maya tone

As regards its prosodic characteristics, Yucatec Maya is exceptional in its language family in being the only Mayan language that has developed lexical tones. It is claimed to have two tones, lexically high and low [4]. A lexical tone appears on syllables containing a long vowel. Syllables with short vowels also exist and these are said to be toneless.
The realization of low and high tones was measured in [4]. In contrast to previous analyses, the high tone is claimed to be realized as a rise in pitch while the low tone remains flat on a low pitch level. The prosodic domain of tone has been claimed to be both the syllable and the mora [3].

2. METHODS

2.1. Subjects

The data reported in this paper was obtained with two native speakers, residents of the village Yaxley (Quintana Roo). Both speakers are native in Yucatec Maya and bilingual (in Spanish). The recordings took place in Berlin during a short visit by both speakers in August 2006.

2.2. Speech Materials

Based on the generalization that patient constituents are often focused in situ in spontaneous discourse (cf. 1.1; [5]), we constructed sentences instantiating this configuration. In particular, we wish to check for prosodic effects on adjectives (embedded within object NPs) comparing a contrastive and a non-contrastive context. The target sentence is given in (3). A syntactic option for focus on the adjective would contain movement into the focus position resulting in a discontinuous NP. Since split NPs are highly marked, the expression in (3) corresponds to a “natural” answer in the presented contexts. This sentence was read by the speaker in the contrastive context (4) or in the non-contrastive context (5). Notice that the target word is not sentence final, contrary to [3].

(3) Target sentence

\begin{align*}
\text{t-in} & \quad \text{w-il-ah} \quad \text{hun-kául} \quad \text{che’ kóom} \\
\text{PFV-1.SG} & \quad \text{0-see-CMPLone-CL-PLANT} \quad \text{tree} \quad \text{short} \\
\text{ich-e} & \quad \text{kóol-o’}. \\
\text{in-DEF} & \quad \text{milpa-d2} \\
& \quad \text{‘I saw a short tree in the milpa.’}
\end{align*}

(4) Contrastive context

\begin{align*}
\text{t-in} & \quad \text{w-il-ah} \quad \text{hun-kául} \quad \text{che’ chowak} \\
\text{PFV-1.SG} & \quad \text{0-see-CMPLone-CL-PLANT} \quad \text{tree} \quad \text{long} \\
\text{ich-e} & \quad \text{kóol-o’}. \\
\text{in-DEF} & \quad \text{milpa-d2} \\
& \quad \text{‘I saw a long tree in the milpa.’}
\end{align*}

(5) Non-contrastive context

\begin{align*}
\text{ba’x} & \quad \text{t-a} \quad \text{w-il-ah} \quad \text{ich-e kóol-o’}? \\
\text{what} & \quad \text{PFV-2.SG} \quad \text{0-see-CMPL.in-DEF} \quad \text{milpa-d2} \\
& \quad \text{‘What did you see in the milpa?’}
\end{align*}

To get evidence for possible interactions between contrast and the different lexical tones, we constructed items for six monosyllabic adjectives, i.e. two for each lexical tone: H, L and N (Tab. 1). Each token was recorded eight times per speaker (in separate sessions), yet some instances had to be excluded for analysis (see n of valid tokens in Tab. 1). The recording sessions contained a number of tasks which were used as fillers for each another.

<table>
<thead>
<tr>
<th>Tone</th>
<th>Carrier Word</th>
<th>n (contrastive / non-contrastive)</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>chup ‘full’</td>
<td>uts ‘beautiful’ 34 / 27</td>
</tr>
<tr>
<td>H</td>
<td>kóom ‘short’</td>
<td>ch’óop ‘blind’ 35 / 27</td>
</tr>
<tr>
<td>L</td>
<td>áal ‘heavy’</td>
<td>bóox ‘black’ 27 / 24</td>
</tr>
</tbody>
</table>

2.3. Analyses

The data were processed in Praat [6] with a 22.05 kHz sampling rate and a 16 bit resolution. The labeling was done at word level including the beginning and end of the sentence. Since the target words were monosyllables, word level labeling equates to the level of syllable in our case. The duration as well as the F0 analysis were conducted semi-automatically using a Praat script. Duration measurements were obtained from the word level labels; the time of the beginning and end of the sentence as well as the target word and its preceding head noun were stored. Time-normalization was done by relating the duration of the target word to the duration of the whole sentence (\(\Delta t_{\text{Adj}}/\Delta t_s\)) and in relation to the duration of the head noun (\(\Delta t_{\text{Adj}}/\Delta t_{\text{NP}}\)).

F0 was extracted using a Hanning window of 0.4 seconds length with a default 10 ms analysis frame. The analysis script allowed for marking of the sonorant part of the target word, which is not equivalent to the duration of the target word in all cases (cf. Table 1). Obvious F0 errors were corrected by hand and F0 was smoothed by 10 Hz. Within the sonorant part two measurements were made: first the pitch of the tonal target was extracted (H, L and N); second, ten points of F0 in equal distance in relation to the duration of the sonorant part of the target word were extracted resulting in time-normalized F0 time courses (pitch normalized according to a speaker’s range [11]).

In the following analyses of pitch and duration we compare the means of the obtained values accompanied by their confidence intervals. We assume that a difference between two means \(x_1\) and \(x_2\) is such that \(x_1 > x_2\) is not significant, if either the higher bound of \(x_2\) is higher than \(x_1\) or the lower bound of \(x_1\) is lower than \(x_2\).
3. RESULTS

3.1. Tonal targets

Figure 1 displays a comparison of means of pitch on the two tonal targets (H and L) as well as the neutral tone syllable (N), for contrastive (black) and non-contrastive (grey) items. As can be seen, only the high tone comparisons yields a significant difference. Interestingly, it is the non-contrastive version that is significantly higher than the contrast version.

![Figure 1: Normalized Pitch values of target items reflecting their tonal targets in comparison between contrastive (black) and non-contrastive (grey) focus.](image)

3.2. Time course of F0

In this part we present the same comparison between contrastive and non-contrastive items in terms of the F0 time course over the target word. Figures 2 to 4 present ten measuring points of the sonorant part of the corresponding syllables for high, low and no-tone syllables respectively.

First, comparing the realization of the three tones, we can confirm [4] that the high tone is realized with a rise in pitch. Both the low and no-toned syllables show a fall in pitch, the latter being realized in a higher register than the former.

Second, comparing the contrastive version with the non-contrastive version in each figure we observe no significant difference between the two except for the high tone. In non-contrastive items the high tone is significantly higher throughout the whole syllable compared to the contrastive version.

![Figure 2: Time course of F0 during the target word, normalized pitch and time normalized. Comparison of lexically high toned words in contrastive (black) and non-contrastive (grey) contexts.](image)

![Figure 3: Time course of F0 during the target word, normalized pitch and time normalized. Comparison of lexically low toned words in contrastive (black) and non-contrastive (grey) contexts.](image)

![Figure 4: Time course of F0 during the target word, normalized pitch and time normalized. Comparison of lexically no-toned words in contrastive (black) and non-contrastive (grey) contexts.](image)

3.3. Duration

Previous studies have shown that contrast may have an effect on the duration of focused elements, both in intonation languages such as German [1] or tone languages such as Chinese [10]. According to such observations, the contrastive version in our experiment was expected to show longer duration than the non-contrastive one.

However, a comparison of the average duration of the adjectives in the contrastive and non-contrastive conditions reveals that this expectation does not hold for Yucatec Maya. Figure 5 presents the means of the $\Delta t_{\text{Adj}}/\Delta t_{\text{NP}}$ ratios calculated per
speaker and item separately. The differences in duration are randomly distributed in the target items. A comparison across items reveals no difference between contrastive and non-contrastive (means of ratios in both cases: 0.508), which is not in accordance with our prediction and which is statistically not significant.

**Figure 5:** Means of ratios ($\Delta t_{Adj}/\Delta t_{NP}$) of duration values in contrastive and non-contrastive contexts; $\Delta t_{Adj}$: adjective duration, $\Delta t_{NP}$: NP duration.

```

4. DISCUSSION
```

This paper is about the prosodic realization of contrastive focus in Yucatec Maya. We compared the prosodic realization of sentences containing an in situ contrasted adjective with the realization of the same sentences in a non-contrastive context. The 174 sentences realized by two native speakers were analysed in terms of duration, tonal target pitch values, and time-normalized pitch course over the target word. In contrast to [3], target words were sentence medial to avoid sentence final effects.

None of the prosodic correlates showed a significant difference between the contrastive and corresponding non-contrastive version except the high tone. It is however the non-contrastive version that is realized higher in contrast to studies that report certain prosodic effects of focus such as H-raising [7]. Phrasing might cause this result in our data since both speakers insert a pause after the target item in contrastive contexts; a phrase boundary lowers the pitch of a high tone [4]. From this we conclude for the present analysis that contrast in Yucatec Maya is not expressed by tonal means (see also [2], [4]). Tone languages need not necessarily to show tonal reflexes of focus such as pitch accents since the primary correlate, F0, is used for expressing differences in lexical tone.

Our findings are in line with previous studies that show that focus is not prosodically expressed ([2], [4]). Since Yucatec Maya has a variety of morphological and syntactic means to express information structure, the above observation might lead to the conclusion that prosody is not necessary as an additional cue to express focus. However, we examined sentences where focus is not expressed by means of syntax or morphology, and where, moreover, a syntactic expression would result in a highly marked structure (split NP) which is rarely used in spontaneous discourse. The results of our study suggest that in the unmarked syntactic configuration there is no effect of contrast at all – neither in prosody nor in syntax.

**Glosses:**

A=person clitic class A; B=person clitic class B; Cl=noun class; CMPL=compleitive aspect; Def=definite; PFV=perfective aspect; SG=singular; SUBJ=subjunctive; TRR=transitivizer.

**Acknowledgements:**

This paper is part of the projects D2 & D5 of the SFB 632 on Information Structure at Potsdam University, financed by the DFG. Thanks to Sam Hellmuth, Christian Lehmann & Bernadett Smolibocki for their individual contributions to this paper.

**5. REFERENCES**


Encoding Information Structure in Yucatec Maya:
On the Interplay of Prosody and Syntax*

Frank Kügler¹, Stavros Skopeteas¹ and Elisabeth Verhoeven²

¹University of Potsdam, ²University of Bremen

The aim of this paper is to outline the means for encoding information structure in Yucatec Maya. Yucatec Maya is a tone language, displaying a three-fold opposition in the tonal realization of syllables. From the morpho-syntactic point of view, the grammar of Yucatec Maya contains morphological (topic affixes, morphological marking of out-of-focus predicates) and syntactic (designated positions) means to uniquely specify syntactic constructions for their information structure. After a descriptive overview of these phenomena, we present experimental evidence which reveals the impact of the non-availability of prosodic alternatives on the choice of syntactic constructions in language production.

Key words: cleft constructions, lexical tone, topic affixes, verb-initial language

1 Preliminaries

Yucatec Maya is a Mayan language spoken by a population of 700,000 speakers (following the 1990 census) at the Yucatecan peninsula (Mexico). As in most other Mayan languages, the canonical order in Yucatec Maya is verb-initial (see England 1991; Norman & Campbell 1978).

With respect to information structure, Yucatec Maya shares with other Mayan languages that preverbal placement of constituents is unambiguously associated with the particular pragmatic functions ‘topic’ and ‘focus’. The topic constituent is placed clause initially and is followed by a deictic suffix. Focus is

* We would like to thank Sam Hellmuth and Ruben van de Vijver for their valuable comments on this paper.
encoded through cleft constructions that place the focused constituent in the immediately preverbal position.

As regards its prosodic characteristics, Yucatec Maya is exceptional in its language family in being the only Mayan language that has developed lexical tones. However, lexical tone and intonation neither interact in the expression of topic nor of focus (Kügler & Skopeteas 2006).

The aim of this paper is to outline the means of encoding information structure in Yucatec Maya. Section 2 presents the morpho-syntactic devices that Yucatec Maya uses for the encoding of topic and focus. Section 3 gives an outline of the tonal characteristics of Yucatec Maya, presenting the prosodic realization of lexical tones and examining the availability of tonal reflexes of information structure. In Section 4, we present the results of a production experiment and we discuss the impact of the grammatical and prosodic properties of Yucatec Maya as outlined in sections 2 and 3 on the choice of grammatical constructions in language production. Section 5 summarizes the main results of this work.

The data were collected in December 2004 in the village of Yaxley (Quintana Roo, Mexico). The subjects that participated in all reported experiments were native speakers of Yucatec Maya and bilingual in Spanish, but exclusively use Maya in their everyday communication within the community.

2 Morpho-Syntactic Encoding of Information Structure

As mentioned in Section 1, verb initial sentences are considered to be canonical based on the criterion of structural markedness: constructions with preverbal arguments are morphologically marked, hence the verb initial order is the

---

1 See Fisher (1976) for discussion of tonogenesis in the Yucatecan branch of Mayan languages.
canonical one (see Durbin & Ojeda 1978). Both VSO and VOS are possible orders in the language, but VOS, which is exemplified in (1), is considered to be pragmatically neutral.²

(1) T-u hàant-ah òon Pedro.
PFV-A.3 eat:TRR-CMPL avocado Pedro
‘Pedro ate avocado.’

Though sentences with two postverbal arguments such as (1) qualify as canonical sentences in Yucatec Maya based on the criterion of morphological markedness, this order only very rarely occurs in corpora (1% in a corpus of 200 clauses, see Skopeteas & Verhoeven 2005). This is a consequence of the fact that verb-initial sentences are thetic, and thetic sentences only rarely occur in discourse. The most frequent sentence type with two lexically realized arguments in a corpus is generally a categorical sentence with a topicalized agent (see (2)).

Topicalized and (narrowly) focused constituents are placed preverbally, while a postverbal constituent be part of a broad focused part of the clause (as, e.g., the patient òon in (2)). A topicalized constituent occurs clause initially and is obligatorily right-bounded by a deictic suffix (a’ ‘D1’: deixis to the 1st person; o’ ‘D2’: deixis to the 2nd person; e’ ‘D3’: contextually given referent).³

(2) Pedro-e’ t-u hàant-ah òon.
Pedro-D3 PFV-A.3 eat:TRR-CMPL avocado
‘As for Pedro, he ate avocado.’

---
³ See also Bohnemeyer (1998) and Lehmann (1990) about topicalization in Yucatec Maya.
Narrow focus is assigned by the displacement of an argument in the preverbal position (cf. (3)). Focus on the agent of a transitive verb triggers a special ‘out of focus’ form of the verb (cf. (4)): the aspect auxiliary is dropped together with the cross-reference clitic for the agent. In the perfective aspect, the extrafocal verb bears the zero form subjunctive marker in non-clause-final position (Bricker 1979, Lehmann 1990). The constructions in (3) and (4) are cleft sentences. The main clause only contains a noun phrase which in Yucatec Maya as in many other Mesoamerican languages may constitute an independent nominal clause. Verb and postverbal argument form a relative clause (relative clauses in Yucatec Maya are not introduced through a relative pronoun). The analysis of these constructions as cleft sentences explains the occurrence of the verb form in (4): This verb form does not occur elsewhere in main clauses, and it is this verb form that is used in relative clauses that are headed by an agent NP.

(3) òon t-u hàant-ah Pedro.
   avocado PFV-A.3 eat:TRR-CMPL Pedro
   ‘It was (an) avocado that Pedro ate.’

(4) Pedro hàant òon.
    Pedro eat:TRR(SUBJ) avocado
    ‘It was Pedro who ate (an) avocado.’

3 Prosody: Tone and Intonation in Yucatec Maya

In this section, we attempt to show that information structural categories such as topic or focus are not expressed by means of post-lexical tones (intonation) in Yucatec Maya. To show this, we introduce the inventory of lexical tones of Yucatec Maya as well as their phonetic realization in section 3.3. Based on these observations, in section 3.4 we analyse target words bearing lexical tones in different syntactic positions that encode distinct information status. Comparing
the realization of lexical tones on target words occurring in broad and narrow focus as well as in topic position, we observe no further tonal effects that might arise due to intonation such as focal tone insertion as in Swedish (Bruce 1977) or Basque (Gussenhoven 2004). The following section (3.1) reviews the literature on the tone system of Yucatec Maya. Section 3.2 then introduces basic methodological issues of the production experiment as well as the data analysis.

3.1 Tone in Yucatec Maya

There is controversy in the literature regarding the tonal system of modern Yucatec Maya. All investigations agree that long vowels are obligatory tone bearing units and display an opposition between a high tone and a low tone. Short vowels are treated as contrasting two levels of pitch in Pike (1946), or as instantiating a third tone termed as “neutral” in Fisher (1976), or as having no tone in Blair & Vermont-Salas (1965). The tonal distinction as well as the distinction between long and short vowels is shown to be contrastive: luk’ul ‘goes away’ - liúk’ul ‘swallow’ - liúk ‘mud’ (examples from Lehmann 1990; see also Blair & Vermont-Salas 1965 and Pike 1946).

As for the realization of tones, authors agree that the low tone is realized as a level tone (Blair & Vermont-Salas 1965, Pike 1946, Straight 1976). Concerning the lexical high tone, three different realizations have been claimed: (i) rising (Blair & Vermont-Salas 1965, Straight 1976) (ii) falling (Fisher 1976), and (iii) falling from high or high level (Pike 1946). However, Fisher (1976) shows that the falling realization occurs in monosyllabic words while in the first syllable of disyllabic words the lexical tone is realized as a rise. None of these investigations argues that the different realizations of a high tone are contrastive at the lexical level.
3.2 Procedure of the production experiment and data analysis

The speech data for the analysis reported in this section were recorded during the same field period as all the other data reported in this paper. Data elicitation took the form of a production experiment with an experimental setup that allows for separating lexical and post-lexical tones. The general procedure is inspired by the work of Bruce (1977) on the tonal aspects of Swedish word accents. Three distinct sentence structures served for the elicitation of target words in broad and narrow focus (post- and preverbal position, respectively), and in topic position (cf. section 2). The structures are listed in (5), where (5a) evokes broad focus in a sentence with the target word as a single argument of the existential verb, (5b) narrow focus in a sentence with the target word in the focus position, and (5c) topic in a sentence with the target word in the topic position. In all sentences the target words are non-initial and non-final, in order to avoid interactions with sentence initial reset or sentence-final lowering. The target words were chosen from the YUCLEX database (Lehmann s.d.), in order to consider instances of all possible tonal patterns (see Table 1).

(5) a. Broad focus construction
   yàan hun-túul ___ ichil le nah-o’.
   EXIST one-CL.AN ___ in DEF house-D2
   ‘There is a ___ in the house.’

   b. Narrow focus construction
   ho’lyak-e’, ___ hàant-ik le òon-o’.
   yesterday-D3 ___ eat:TRR-INCMP DEF avocado-D2
   ‘Yesterday, it was ___ who ate the avocado.’

---

4 In this article, we particularly discuss the realizations of the minimal pair miis ‘cat’ and miis ‘broom’.
c. Topic construction

ku ts’o’kol-e’ le ___-e’ h bin-ih.

‘Afterwards, what the ___ concerns, (s)he went away.’

Table 1: Tonal patterns in lexical items. (N = neutral; L = low; H = high; grave accent indicates low tone, acute accent high tone).

<table>
<thead>
<tr>
<th>tonal pattern</th>
<th>lexical item</th>
<th>translation</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>am</td>
<td>spider</td>
</tr>
<tr>
<td>L</td>
<td>lòol</td>
<td>flower</td>
</tr>
<tr>
<td>L</td>
<td>miis</td>
<td>cat</td>
</tr>
<tr>
<td>H</td>
<td>miis</td>
<td>broom</td>
</tr>
<tr>
<td>H</td>
<td>láal</td>
<td>stinging nettle</td>
</tr>
<tr>
<td>N-N</td>
<td>ahaw</td>
<td>chief</td>
</tr>
<tr>
<td>N-L</td>
<td>konküum</td>
<td>pot seller</td>
</tr>
<tr>
<td>N-H</td>
<td>konchúuk</td>
<td>shoe seller</td>
</tr>
<tr>
<td>L-N</td>
<td>yúuyum</td>
<td>bird</td>
</tr>
<tr>
<td>L-H</td>
<td>kòolnáal</td>
<td>farmer</td>
</tr>
<tr>
<td>L-L</td>
<td>xtiuuxkiuuts</td>
<td>pheasant</td>
</tr>
<tr>
<td>H-N</td>
<td>yáalam</td>
<td>fawn</td>
</tr>
<tr>
<td>H-L</td>
<td>óochkàan</td>
<td>snake</td>
</tr>
<tr>
<td>H-H</td>
<td>tôokchüuk</td>
<td>coal merchant</td>
</tr>
</tbody>
</table>

The speech data were elicited by means of question-answer pairs. Since most Yucatec Mayan speakers are not trained in reading Mayan orthography, we had to present our stimuli orally. The carrier sentences with target items as given in Table 1 were thus read by a native speaker before running the experimental sessions. The pitch contour of each provided sentence, however, was reduced to a flat level pitch in order to eliminate all linguistic information that is encoded by pitch. In the experimental sessions, informants heard the resynthesized stimuli. The informants' task, then, was to answer a generic question by repeating the text they had just heard before. All recordings were made on a DAT recorder (SONY 100) using head-mounted microphones. For the
manipulation of the test sentences and for pitch analysis we used Praat (Boersma & Weenink 2006).

In total, twelve (male and female) speakers have been recorded. However, all twelve speakers did not produce sentences with all test items. The individual time-normalized measurements are based on two to six speakers.

The pitch analysis has been made using a hanning window of 0.4 seconds length with a default 10 ms analysis frame. The pitch contour has been smoothed using the Praat smoothing algorithm (frequency band 10 Hz) to diminish microprosodic perturbations. Following Xu (1999) the pitch tracks were time-normalized with ten measuring points during the voiced part of each of the labeled intervals. The time-normalized plots reported below thus only refer to the voiced parts of the words leaving voiceless parts aside. The F0-values measured in Hertz were converted to semitones5 to normalize across the physiological differences of male and female voices.

3.3 The phonetic realisation of lexical tones

In this section we provide an overview of the realization of lexical tones in Yucatec Maya. It has been claimed that Yucatec Maya exhibits a tonal distinction between high and low tones with additional toneless syllables (e.g. Pike, 1946). We will show the realisation of low and high tones below.

On the basis of a first inspection of empirical data gained through our production experiment, in Kügler & Skopeteas (2006) we identified a lexical low and a lexical high tone. The data presented here show systematically

5 The conversion from Hertz into semitones is made according to the equation below with an arbitrarily chosen reference of 100 Hz (e.g. Reetz 1999):

\[ f_{(st)} = 12 \log_2 \left( \frac{f_{(Hz)}}{100 \text{ Hz}} \right) \]

See also Nolan (2003) who has convincingly demonstrated that the semitone scale fits best the intonational equivalence scale; see Ladd (1996:260ff) for the notion of semitones with respect to pitch range.
analysed and normalized results. While the low tone is realized as a low level tone (see Fig. 1, right panel) which is in accordance with the previous accounts discussed above, we find evidence for the high tone being realized as a rise in pitch (see Fig 2), thus supporting the analysis of Blair & Vermont-Salas (1965), Straight (1976), and partly that of Fisher (1976). The difference in the observed contours in Fig. 1 is due to a difference in tonal structure. In the broad focus condition (left panel of Fig. 1) a high tone on the indefinite marker *huntúul* precedes the lexical low tone of the target word, whereas a toneless syllable precedes the target syllable in the topic condition.

Fig. 1: Target word *miis* ‘cat’ with lexical low tone; in postverbal position (broad focus and canonical word order, five speakers) in the left panel, and in preverbal topic position in the right panel (five speakers).

Fig. 2: Target word *miis* ‘broom’ with lexical high tone in postverbal position (broad focus and canonical word order, four speakers).
For syllables containing long vowels this tonal distinction is obligatory, whereas syllables containing short vowels are empty TBUs.

3.4 Tonal effects of information structure

In this section, we argue that information structure does not affect the realization of lexical tones in Yucatec Maya, i.e. the language does not encode a particular focus structure by means of intonational tones. We observe no interaction of lexical tones and post-lexical tones (intonation) meaning that Yucatec Maya does not employ additional pitch accents to express topic or focus. Properties of Yucatec Mayan intonation are dealt with in Blair & Vermont-Salas (1965) who offers a detailed annotation of intonational contours made for didactic purposes. Furthermore, Straight (1976) gives an inventory of rules that predict different realizations of the lexical tones in several tonal environments.

As already shown in Section 2, a crucial aspect of the Yucatec Mayan grammar is that the syntactic realization of the arguments is determined by information structure. As a result, it is not possible to examine the prosodic effects of information structure independently of syntax, i.e. it is not possible to design minimal pairs of identical carrier sentences that will be produced in contexts that induce distinct information structures. With this constraint in mind, the question of prosodic effects of information structure in Maya may be inspected by using the same lexical unit in different information structural – but necessarily also syntactic – positions (cf. (5) above).

In Kügler & Skopeteas (2006) we investigated the interaction between lexical tones and intonation in Yucatec Maya on the basis of a first inspection of the data, and concluded that there appears to be no tonal means for the expression of focus or topic. As illustrated in Section 2, narrow focused constituents appear preverbally (cf. (5b) above). If a word containing a lexical pre-specified tone occurs in the focus position, the underlying shape of the tone
as described in Section 3.3 remains preserved. Thus, we observe no interaction of lexical tone and intonation (in the form of particular pitch accents) for the expression of focus. The data presented here are calculated means of six speakers. Consider the pitch track of the monosyllabic target word *miis* ‘broom’ with lexical high tone in narrow focus position in the left panel of Fig. 3. The target word is realized with the rise in pitch identical to the high tone rise established in Section 3.3. There appears no further tonal event that might be analyzed as a pitch accent indicating focus tonally. If we compare the narrow focus realization of a target word containing a lexical high tone with a realization in broad focus (postverbally, cf. Fig. 2 above) or in topic position (preverbally as in the narrow focus condition, cf. the right panel of Fig. 3), we observe the same tonal pattern, i.e., a rise in pitch on the target word. Thus, we conclude that information structural components such as topic, narrow and broad focus are not expressed by means of post-lexical tones (pitch accents) as is the case in languages such as Basque or Swedish (cf. Gussenhoven 2004).

**Fig. 3:** Target word *miis* ‘broom’ with lexical high tone in preverbal position, normalized across six speakers; the left panel shows narrow focus, and the right panel topic position.
3.5 Summary

Concerning the prosodic properties of Yucatec Maya, we have shown that a lexical low tone is realized as a low level pitch, whereas a lexical high tone is realized as a rise in pitch (cf. also Kügler & Skopeteas 2006). Further, the realization of the lexical tones is not affected by information structure, i.e. information structure appears to induce no qualitative effects on the realization of lexical tones. Focus is mainly expressed by means of syntax, which is explored in more detail in the next section.

4 Encoding Information Structure in Spontaneous Production

We have seen in Section 2 that Yucatec Maya displays a number of morpho-syntactic devices for the encoding of topic and focus. We have shown in Section 3 that the language does not employ tonal means for the encoding of information structure. In this Section, we present data obtained through a production experiment in which spontaneous responses to different question types were elicited. This data shows the impact of the mentioned structural and prosodic properties on the choice of a given grammatical construction in particular discourse conditions.

The experiment discussed in this paper was developed within the project D2 “Typology of Information Structure” (part of the SFB 632 “Information Structure”) and is part of the Questionnaire on Information Structure (Skopeteas et al. 2006)\(^6\).

\(^6\) See Skopeteas et al. (2006:119ff.) for a full documentation of the experimental procedure and material.
**Experimental procedure**

Subjects were shown a sheet of paper that contained four pictures. All pictures presented situations in which two entities were involved in events which are typically encoded by transitive verbs \( x \) hits \( y \), \( x \) kicks \( y \), \( x \) carries \( y \), \( x \) cuts \( y \), etc.). The subjects were given one minute to observe what happens in the pictures; after that, the pictures were taken away and four questions relating to the pictures were played from a laptop. The questions were pre-recorded with two native speakers on a DAT recorder (SONY 100) and digitized at a sampling frequency of 22050 Hz. Subjects were instructed to listen to the questions and give a spontaneous answer. This experiment was part of a longer session (about 40 min.) that contained pseudo-randomized tasks from six different production experiments that were used as fillers for each other.

Sixteen native speakers of Yucatec Maya participated in this experiment. Their responses were recorded using head-mounted microphones on the same DAT recorder.

**Experimental conditions**

For the purposes of this paper, we will consider four of the eight conditions of this experiment. Two factors are instantiated in the four conditions:

(a) solicited argument: agent or patient;

(b) relation of the (intended) answer to the question: completive (i.e., filling a gap in the presupposed information) vs. corrective (i.e., replacing a part of the presupposed information).

The combination of these factors results in four conditions. The questions establishing these conditions are exemplified in (6).
The conditions exemplified in (6) were factorially implemented in 16 items presenting different events, all involving two participants. Each subject was confronted with each item once and with each experimental condition twice. Thus the experimental procedure resulted in a corpus of $(16 \times 2 =) 32$ answers per condition, which are discussed in the following Subsection.

**Results**

In accordance with the syntactic properties of the language which have been presented in Section 2, all question types presented in (6) elicited focus constructions to some extent. In the following examples, the argument which is
solicited through the question is placed preverbally and the argument which is part of the background of the question is placed postverbally.

(7)  a. A-focus
Q= Who is looking at the girl?
  hun-túul xibpal pak-t-ik
  one-CL.AN man:child see-TRR-INCMPL
  le x-ch’úuppal-o’.
  DEF F-woman:child-D2
  ‘It is a boy that is looking at the girl.’

b. P-focus
Q= What is the man kicking?
  hun-p’éel esten... k’áanche’ k-u
  one-CL.INAN HESIT chair IPFV-A.3
  kóochek’-t-ik le xìib-o’.
  kick:foot-TRR-INCMPL DEF man-D2
  ‘It is a chair that the man is kicking.’

The solicited information may also occur postverbally, as in the following example. The argument which is part of the background of the question is placed in the topic position.

(8)  Q= Is the man kicking a table?
  le xìib-o’ túun kóochek’-t-ik
  DEF man-D2 PROG:A.3 kick:foot-TRR-INCMPL
  hun-p’éel silla
  one-CL.INAN chair
  ‘The man is kicking a chair.’

Answers with two preverbal arguments also occur, but only in the conditions in which the agent is a topic and the patient is in focus (and not vice versa).
(9)  Q= Is the woman hitting a flower?
    ma’, le x-ch’úup-o’ hun-p’éel k’áax k-u
    NEG DEF F-woman-D2 one-CL.INAN wood IPFV-A.3
    lox-ik.
    hit-INCMPL
    ‘No, the girl hits a piece of wood.’

Since the subjects were instructed to give a spontaneous answer to the recorded
questions, the results contain also elliptical sentences that do not allow insights
into the function of sentential positions.

(10)  Q= Who is carrying the pot?
    hun-túul máak.
    one-CL.AN man
    ‘A man.’

Argument ellipsis is attested, too. In these answers, the focused argument and
the verb are realized and the argument which is part of the question’s back-
ground is elided. There are two possible realizations of focused arguments in
these sentences, either in the preverbal focus position (11b) or postverbally
(11a).

(11)  a. Postverbal realization
    Q= What is the man pulling?
    túun kóol-ik hun-p’éel mesa
    PROG:A.3 pull-INCMPL one-CL.INAN table
    ‘He is pulling a table.’

    b. Preverbal realization
    Q= What is the man carrying?
    hun-túul x-ch’úup k-u bis-ik
    one-CL.AN F-woman IPFV-A.3 carry-INCMPL
    ‘It is a woman that he’s carrying.’

Fig. 4 and Fig. 5 show the results gained in the four experimental conditions.
Note that the figures only contain those answers that (a) do not imply a
misinterpretation of the stimulus and that (b) do not display verb ellipsis. Three types of answers are distinguished in Fig. 4: focus constructions as exemplified in (7), postverbal placement of the solicited information as illustrated in (8) and “other”. The cases classified as “other” contain pseudo-clefts or complex sentences with a dislocated argument.

Fig. 4: Encoding the argument which is solicited through the question (“Ag”: agent in focus; “Pat”: patient in focus; “Compl”: completive answer; “Corr”: corrective answer).

Fig. 4 shows an asymmetry in the encoding of agents and patients, when solicited through the questions. Agents are almost always placed in the preverbal focus position, while patients may occur in the postverbal position, too. This result reveals that patients may also be focused in situ. Furthermore we can observe descriptively in Fig. 4 that the type of question (i.e., corrective vs. completive) does not have an impact.

An argument which is part of the background of the question can be encoded either as a topic (see (8) and (9)), or postverbally (see (7)), or is elided (see (11)). The observed occurrence of these options is presented in Figure 5: in agent questions the background argument is the patient, and in patient questions the background argument is the agent.
Figure 5 shows that an asymmetry holds for the topicalization of the arguments, too: an agent is placed in the topic position when it is background information (see patient questions), while a patient which is the background information is always encoded postverbally. As already observed with respect to the encoding of the solicited argument, the difference between corrective and completive answers does not crucially affect the encoding of the background argument in Yucatec Maya.

The common denominator between the two observed asymmetries is a general preference for Agent first orders. In sentences with two third person nominals, one of which is topicalized, the construction is ambiguous due to the lack of case marking. For these sentences, there is a strong preference to interpret the topicalized argument as an agent which probably results in the asymmetry presented in Fig. 5.

5 Summary

We demonstrated in Section 2 that Yucatec Maya provides unambiguous syntactic means for the encoding of information structure. In Section 3, we
illustrated that information structure appears to induce no qualitative effects on the realization of lexical tones. The consequence of these observations is that speakers will have to choose a syntactic device in order to encode the information structure of arguments. This hypothesis was confirmed by a production experiment that we presented in Section 4: in all question types, speakers produced a high amount of syntactic constructions that focused the solicited argument. The data obtained by this experiment also revealed an asymmetry between focused patients and focused agents: while agent questions almost always triggered agent focus, patient questions only triggered patient focus in half of the answers. In the other half, the patient – though it is the solicited argument – remains in situ. The data from topicalization revealed a reverse asymmetry; the agent is the preferred topic while the patient does not occur in topic position. Both asymmetries are attributed here to a general preference for agent first orders in Yucatec Maya.

6 Glosses

A cross-reference marker, set A
B cross-reference marker, set B
CL classifier
AN animate
INAN inanimate
CMPL completive
D deictic
DEF definite
EXIST existential
F feminine
HESIT hesitative
INCMPL incompletive
INT interrogative
IPFV imperfective
NEG negative
PFV perfective
PROG progressive
SUBJ subjunctive
TRR transitivizer

References


Lehmann, C. s.d., YUCLEX. Erfurt: Univ. of Erfurt.


Section IV

Akan
On the prosodic expression of pragmatic prominence – The case of pitch register lowering in Akan

Frank Kügler and Susanne Genzel
University of Potsdam, Department of Linguistics

Running head: Pitch register lowering in Akan

Acknowledgements: This paper is part of the research project D5 “Tonal structure and information structure” within the Collaborative Research Centre on Information Structure at Potsdam University which is funded by the German research foundation DFG (SFB 632). Parts of this paper have been presented at the University of Ghana, January 2009, at the annual meeting of Phonetik und Phonologie-5 in Cologne, March 2009, and at the workshop on Intonation and Meaning, Barcelona, September 2009. We thank all participants for valuable discussion. We would like to thank, in alphabetical order, a number of people who contributed to this work: Dinah Baer-Henney, Afua Blay, Daniel Büring, Heiner Drenhaus, Kofi Dorvlo, Laura Downing, Ines Fiedler, Sophia Jähnigen, Annie Riallant, Kofi Saah, Bernadett Smolibocki, Anne Schwarz, Hubert Truckenbrodt, Ruben van de Vijver, and Malte Zimmermann. This paper has benefited from the detailed and thorough thoughts of two reviewers, Bettina Braun and Bert Remijen as well as the Associate Editor Marc Swerts.

Corresponding author: Frank Kügler
University of Potsdam
Department of Linguistics
Karl-Liebknecht-Str. 24/25,
D – 14476 Potsdam, Germany
Email: kuegler@uni-potsdam.de
ABSTRACT
This article presents data from three production experiments investigating the prosodic means of encoding information structure in Akan, a tone language that belongs to the Kwa branch of the Niger-Congo family, spoken in Ghana. Information structure was elicited via context questions that put target words either in wide, informational, or corrective focus, or in one of the experiments also in pre-focal or post-focal position rendering it as given. The prosodic parameters F0 and duration were measured on the target words. Duration is not consistently affected by information structure, but contrary to the prediction that High (H) and Low (L) tones are raised in ex situ (fronted) focus constructions we found a significantly lower realisation of both H and L tones under corrective focus in ex situ and in situ focus constructions. Givenness does not seem to be marked prosodically. The data suggests that prominence is expressed prosodically by means of a deviation from an unmarked prosodic structure. Results are thus contradicting the view of the effort code that predicts a positive correlation of more effort resulting in higher F0 targets.
INTRODUCTION

This paper addresses two central questions with respect to tone languages. The first one deals with the issue of intonation in a tone language which we will illustrate with data from Akan. The phonetic correlate of tone is F0, hence a Low tone (L) is expressed by means of relatively lower F0 level than a High tone (H). The most salient phonetic correlate of intonation is also F0, that is to say pitch accents and boundary tones are expressed by means of local F0 minima and maxima (Pierrehumbert, 1980; Ladd, 1996; Gussenhoven, 2004). Given the fact that the phonetic correlate F0 is the same for tone and intonation the critical question is thus, how much intonation do tone languages use? According to Ladd (1996, p. 6) intonation expresses post-lexical pragmatic meanings. These refer basically to two aspects, highlighting of information and structuring information. In this paper we are focusing on the former function of intonation, in other words the prosodic expression of new and given information within a phrase. We will present a case of pitch register lowering\(^1\) as the prosodic expression of prominence, which has to our knowledge not been reported yet.

The second issue addressed in this paper concerns the universality of the prosodic expression of prominence. According to Gussenhoven (2002, 2004) the theory of the biological codes represents universal functions of language related to intonation. With respect to prominence Gussenhoven argues that the effort code explains the prosodic expression of focus in terms of higher or later F0 maxima. We will interpret the data of the present study on the prosodic expression of prominence in Akan in the light of Gussenhoven’s theory of biological codes.

To assess the range of variation of languages and to critically reflect typological generalisations, which usually are based on well studied languages, it is necessary to gain more data by means of controlled studies on less well examined languages. Akan represents a

---

1 The notion of pitch register is used in a non–uniform way in the linguistic literature (see Rietveld & Vermillion, 2003 for an overview). We assume here that the pitch register sets reference lines relative to which tones are scaled (Clements, 1979).
good candidate for this enterprise for two reasons. First, Akan is a level tone language (cf., Pike, 1948). Controlled data on level-tone languages, in particular from the African continent, is rare (Zerbian, Genzel & Kügler, 2010) while controlled data is much more available for contour tone languages like Chinese (e.g., Xu, 1999; Liu & Xu, 2005; Chen, Wang & Xu, 2009). Second, a large amount of papers deal with aspects of the tonal system, of the information structure, and of the syntax of Akan, which we can base our study on. As yet, however, no study has quantitatively investigated the prosodic expression of information structure in Akan.

In the remainder of the introduction we will present earlier work on Akan tonal phonology, and discuss the issues of intonation in tone languages and intonational universals in more detail. The main finding of the paper is that speakers express intonational function in Akan by means of pitch register lowering, which from a universal perspective contradicts the theory of the biological codes put forward by Gussenhoven (2002, 2004). Contrary to Gussenhoven we will propose that highlighting information in communication is achieved by means of a deviation from a neutral register, and Akan represents a case where speakers lower their voice in order to draw attention of their interlocutors.

**Earlier work on tone in Akan**

This section will present an overview of Akan phonology as it emerges from earlier work. Akan belongs to the Kwa branch of the Niger-Congo family and is spoken by about 8.3 million people in Ghana and the Ivory Coast (Christaller, 1933; Lewis, 2009). Akan consists of several dialects, some of which are more mutually intelligible than others (Schachter & Fromkin, 1968). The dialects differ at the level of segments as well as tones (cf., Cahill, 1985; Dolphyne, 1988; Abakah, 2002, 2005a, b; Abakah & Koranteng, 2006; among others). The three main dialects are Asante Twi (2.8 million speakers), Fante (1.9 million speakers) and Akuapim (0.55 million speakers) (Schachter & Fromkin, 1968, Cahill, 1985; Lewis 2009).
This paper concentrates on Asante Twi which has its cultural centre in and around the city of Kumasi. Asante Twi is the dialect with the majority of speakers, and we will use the cover term Akan throughout the paper.

Akan is a two-tone language distinguishing between L and H tones, that are transcribed in the tradition of the African tone literature and in line with IPA conventions as [`] and [´] respectively (Dolphyne, 1988). All examples throughout the paper are given in Roman script following the orthographic convention laid out in several dictionaries of Akan (e.g., Christaller, 1875, 1933; Mohr, 1909; Berry, 1960; cf., also Dolphyne, 1988). Tone languages are characterized by the way they use F0 to distinguish lexical or grammatical meaning (e.g., Pike, 1948; Hyman, 2001). In (1) the lexical meaning of Akan words changes as a function of differences in lexical tone. In (1a) the disyllabic word carries two lexical H tones whereas in (1b), two lexical L tones, and in (1c) the two syllables carry two distinct lexical tones, L and H. In Akan, tone has grammatical function as well (Dolphyne, 1988). In (2), the second tone of the verb determines the aspect of the verb: the habitual form (2a) is characterized by a H tone, and the stative form (2b) by a L tone. The underlying lexical H tone is post-lexically replaced by a grammatical tone. The grammatical function of tone in Akan plays a more important role than its lexical function (Dolphyne, 1988; Manyah, 2006). Among other things, tones in Akan distinguish verb aspect and tense, and identify argument structures of the verb.

(1)  a. pápá ‘good’

b. òpà ‘fan’

c. pàpà ‘father’

(Dolphyne, 1988, p. 52)

---

2 Akan orthography uses symbols from Roman script as well as additional symbols from the IPA with their corresponding IPA value. For instance, the grapheme <ê> represents a half-open unrounded front vowel (Dolphyne, 1988, p. 7). In total, Akan uses seven vowel, and 16 consonant letters.
The tone bearing unit (TBU) is the syllable (Stewart, 1965; Dolphyne, 1988; Abakah, 2005a, b). Examples are given in (3). According to Dolphyne (1988) Akan distinguishes three syllable types (V, CV, C), which are either open syllables, or single sonorants function as syllabic consonants (cf., also Christaller, 1933, p. XXVIII; Stewart, 1965). Any vowel constitutes a syllable, and in case of two adjacent vowels each of them constitutes its own syllable (3e) (cf., Christaller, 1933, p. XVII; Dolphyne, 1988).

(3) a. ṣ-fā ‘he takes it’  
b. sò-rí ‘hold it’  
c. ñ-sú ‘water’  
d. dà-ń ‘turn it over’  
e. ò-hú-ì ‘he saw it’

The first syllable in (3a) represents an open syllable without an onset, while the second one contains an onset and nucleus. The data in (3b – d) show examples of syllabic consonants either as prefix or as suffix. In (3b) and (3c) the tone associated with that syllable differs from the tone of the following or preceding syllable, while the tone is equal on both syllables in
(3d). Tone bearing prefix consonants such as in (3c) provide evidence that single consonants constitute syllable nuclei. Akan allows no complex onsets as (3c) may suggest (Dolphyne, 1988; Marfo & Yankson, 2008). Tone associates with the rime of the syllable, not with onsets (cf., Yip, 2002).

The distinction of lexically H and L tones is also reflected at the segmental level. According to Manyah (2006), vowels that carry a L tone are significantly shorter than vowels carrying a H tone. The durational difference ranges between 80 to 100 ms. Vowel quality does not differ for different tones (although there is a tendency for H toned vowels being realized more centrally).

According to Christaller (1933) Akan has stress, which is further corroborated by an analysis of speech rhythm in Akan oral praise poetry since Akan rhythm shows more common characteristics with English stress timing (Purvis, 2009; Anderson, 2009). However, to our knowledge, no comprehensive analysis of stress in the sense of Hayes (1995) exists, and the exact phonetic details of the “emphasis put on a syllable” (Christaller, 1933, p. XXVIII) remain unclear. Christaller (1933) at least provides a tentative distribution of stress in Akan: He differentiates between stress on verbs and on nouns. Stress in verbs can fall on the prefix or on the stem. In nouns stress lies either on the first H toned syllable, or on the preceding L toned syllable. In nouns containing only L tones the first stem syllable carries the stress, as in wɔʃa – ‘uncle’. However, assuming H tones correlate with stress may be a misleading perceptual impression. Any analysis of stress in a tone language has to be independently motivated.

### Insert Figure 1 about here
Akan has been classified as a terraced-level tone language (Clements, 1979; Dolphyne, 1988; Abakah, 2000). Welmers (1959) was the first who introduced the term “terraced level languages” for describing languages where the realisation of phonologically identical tones can be quite different depending on where the tones occur in the utterance. According to Clements (1979) terracing languages display a regular process of register shift which affects the F0 realisation of successive tones. The shift of the total pitch register can apply downward (downstep) and/or upward (upstep). An important feature of tone terracing is that there is no limit on the number of register lowerings in a tone group. External limits can be set by lexical, grammatical, and/or phonological factors. Terracing is a process on the level of the pitch register. The terracing property is illustrated in Figure 1, the example is given in (4).

(4) ɓ-bé-kọ  Kùmásé  ánọpá  yí
3SG-FUT-go  Kumase  morning  this

‘He will go to Kumase this morning.’

(Schachter & Fromkin, 1968, p. 105)

Each first H tone in a series of High tones of the utterance in (4) defines a new register line relative to which subsequent tones are scaled. In (4) three distinct pitch register levels exist (cf., horizontal lines in Figure 1).

The terracing effect shown in **Fehler! Verweisquelle konnte nicht gefunden werden.** is analyzed as downstep or downdrift. This phenomenon is also referred to as automatic (downdrift) and non-automatic downstep (downstep). The terms downstep and downdrift are frequently confused in the literature. This is partly due to the fact that

---

3 Clements (1979, p. 537) lists a number of other languages that show tone terracing, which are not limited to sub-Saharan languages such as Niger-Congo (e.g., Yoruba), or Bantu (e.g., Sotho), but also occur in Nilo-Saharan (e.g., Luo), and Chadic languages (e.g., Ga’anda), as well as in some native north American languages such as Acatlán Mixtec.
researches refer to these processes as either phonetic or phonological effects. Another reason is that different research traditions use different terminology, for example African tone linguistics vs. intonational phonology (cf., Clements, 1979; Connell & Ladd, 1990 for an overview). Automatic and non-automatic downstep usually represent a lowering of a H tone due to a preceding L tone. In case of automatic downstep, this L tone is realized phonetically, while it is not in the case of non-automatic downstep. In the latter case, the downstep trigger does not appear on the surface. Based on studies that do not find any phonetic difference between the realisation of automatic and non-automatic downstep (Laniran, 1992 for Igbo; Snider, 1998 for Bimoba, 2007 for Chumburung) we use the term downstep for both types of terracing effects. According to Schachter and Fromkin (1968), Dolphyne (1988), and Abakah (2000, 2002) downstep occurs in Akan in a tonal sequence of H-L-H, where the second H tone is lowered with respect to the first H tone, as illustrated in Fehler! Verweisquelle konnte nicht gefunden werden. 

Tone languages and intonation

From a typological point of view any language is assumed to have intonation (Hockett, 1963; Bolinger, 1962; Gussenhoven, 2004). In tone languages, F0 distinguishes lexical meaning and expresses grammatical relations by means of tone. Given so, it is worthwhile to find out to what extent the phonetic cue F0 may also express post-lexical meanings. Surveying the literature on this issue reveals that in fact tone languages do express post-lexical meanings by means of F0.

Intonation expresses sentence-level meanings such as highlighting information which commonly is known as focus marking. In one group of tone languages focus marking is accompanied by changes in F0 scaling, in other words an enhancement of the tonal register (Xu, 1999; Liu & Xu, 2005; Chen et al., 2009 for Mandarin Chinese, Sinitic; Pan, 2007 for Thai, Tai; Jannedy, 2007 for Vietnamese, Austro-Asiatic; Leben, Inkelas & Cobler, 1989;
Inkelas & Leben, 1990 for Hausa, Chadic; Schwiertz, 2009 for Beaver, Athabaskan). Xu (1999) showed that all four tones in Mandarin Chinese are articulated with expanded pitch register, for example a H tone is raised higher, and a L tone is lowered. In addition to a more enhanced production of tones under focus the tonal register after the focus is suppressed. Also, longer durations of focused words have been reported for Thai (Pan, 2007) and Vietnamese (Jannedy, 2007). As another strategy, the Curaçao dialect of Papiamentu, a Caribbean Creole with lexical tone contrasts, employs a particular focus pitch accent to mark prominence (Remijsen & van Heuven, 2005) which resembles the Swedish focal accent (Bruce, 1977).

A second group of tone languages employs a different strategy to express focus, which refers to prosodic domain structure, that is to say an insertion of a phrase break before or after a focused constituent (Frajzyngier, 1989 for Pero, Chadic; Kidda, 1993 for Tangale, Chadic; Kanerva, 1990; Downing, Mtenje, & Pompino-Marschall, 2004; Downing, 2008 for Chichewa, Bantu; Karlsson, House, Svantesson, & Tayanin, 2007 for Kammu, Mon-Khmer; Schwiertz, 2009 for Beaver, Athabaskan). A phrase break is realized by means of phonetic cues such as a pause, final lengthening (Martin, 1970; Lehiste, 1972; Wightman, Shattuck-Hufnagel, Ostendorf, & Price, 1992), and/or F0 register resetting (e.g., the Bantu language Chichewa: Kanerva, 1990; Downing et al., 2004; Downing, 2008; and some Kwa languages spoken in Côte d'Ivoire, yet not Akan: Leben & Ahoua, 2006). In Kammu, a focus is signalled by means of a H boundary tone (Karlsson et al., 2007).

Contrary to languages of the first two groups, a third group of tone languages does not use prosodic means for the expression of prominence at all. These include the Bantu language Northern Sotho (Zerbian, 2006), the Mayan language Yucatec Maya (Kügler & Skopeteas, 2006, 2007; Kügler, Skopeteas & Verhoeven, 2007; Gussenhoven & Teeuw, 2008), and Navajo which belongs to the Athabaskan language family (McDonough, 2002). Hartmann and Zimmermann (2007) present evidence from production data and perception tests that also
Hausa does not use prosodic means for the encoding of focus. Their results go against the findings of Leben et al. (1989) and Inkelas and Leben (1990).

This review shows that besides tonal distinctions of lexical and/or grammatical functions tone languages do use prosodic means for the expression of post-lexical pragmatic meanings, yet not all tone languages necessarily employ prosodic means for the expression of post-lexical pragmatic meanings.\(^4\) Akan has as yet not been analysed in depth with regard to its intonation and the prosodic expression of focus. Impressionistically, Boadi (1974) reports on tonal raising of both H and L tones on \textit{ex situ} focused words. Thus, we expect Akan to belong to the first group of tone languages that use pitch register in order to express focus.

Another example of intonation in Akan is the expression of sentence mode. Dolphyne (1988, p. 55) reports that interrogatives display a higher overall F0 and that a final H tone is realized with a slightly falling F0. According to Hyman (2001) questions in Akan are further characterized as not having downstep. Yet another intonational phenomenon concerns the change of tones of a whole phrase. Abakah (2005a) gives an example in Akan where a complete main clause receives H tones if a certain temporal adverbial appears in the sentence.

These facts further show that intonation exists in Akan.

\textbf{Intonation and its biological codes}

Gussenhoven (2002, 2004) argues that intonational meaning manifests itself in universal and language-specific aspects. The language-specific aspects are coded in the intonational grammar of each language while the universal part is expressed phonetically across all languages. According to Gussenhoven, three biological codes explain the universal aspects of

\(^4\) This grouping is not a generic property of tone languages. A similar grouping can be achieved for intonation languages which however are not the focus of the present paper. Group one would include languages that express focus with higher or later pitch peaks such as German (see Braun, 2006; Féry & Kügler, 2008; Kügler, 2008; Kügler & Genzel, submitted;) or English (see Bartels & Kingston, 1994; Cooper, Eady, & Mueller, 1985; Eady & Cooper 1986; Eady, Cooper, Kloouda, Mueller, & Lotts, 1986). Group two refers to languages such as Bengali employing a boundary tone as a demarcation of a focus phrase (Hayes & Lahiri, 1991), and group three to languages such as Wolof that do not show any prosodic reflex of focus at all (Riallant & Robert, 2001).
the interpretation of pitch variation, and they are based on physiological properties of the speech apparatus. The frequency code relates dominance relations to pitch height, that is lower pitch is interpreted as more dominant than higher pitch. The effort code relates the amount of energy put into speech production to the amount of emphasis of a particular message, in other words the more effort in production the more emphatic the utterance and the more likely a message comes across. The production code, finally, relates the relative pitch height at the boundaries of utterances to effects of finality, topic initiation, or continuation demands in communication.

To achieve a particular interpretation a speaker can consciously manipulate the code. The expression of emphasis or prominence, hence the manipulation of the effort code, which is most relevant for the present discussion, results in an increase of articulatory precision accompanied with faster vocal cord vibration. In intonation languages the more precise articulation manifests itself in raising of the pitch register. The same effect has been reported for the tone language Mandarin Chinese for H tones (Xu, 1999). Thus, Gussenhoven advocates a correlation that an increase of prominence is achieved by an increase in effort. Gussenhoven further argues that languages have grammaticalised the biological codes, and the effort code represents the expression of focus. However, the theory of biological codes is based mainly on research on intonation languages.

At first sight, the group of tone languages that uses prosodic means for the expression of focus would correspond to the grammaticalised function of the effort code such that they use an expanded pitch register to express focus. Since we expect Akan to belong to that group as well, Akan should also express focus by means of an expansion of pitch register. However, as we will show in this paper, Akan does not behave as expected in the prosodic expression of focus. Contrary to pitch register expansion, in Akan both H and L tones are lowered, which we analyse as a lowering of the pitch register. In the discussion we will argue that these findings contradict the idea of the effort code.
The expression of prominence in Akan

Prominence is understood as highlighting relevant information within an utterance (e.g., Chafe, 1974). In this sense, the information is in focus. A focus of a sentence represents a word or constituent that receives prominence either by means of syntax, morphology, prosody or a combination thereof. In intonation languages such as German or English speakers dominantly use prosodic means to highlight information. This is known as prosodic focus (Ladd, 1980; Gussenhoven, 1984; Ladd, 1996). Semantically, focus defines a set of alternatives from which one element is chosen by the speaker (Rooth, 1985, 1992; Krifka, 2008). If focus highlights new information, parts of a sentence may contain already old information which is usually referred to as given information, defined here as previously mentioned in the discourse (Allerton, 1978; Baumann, 2006).

Prominence in Akan is achieved by means of different linguistic strategies. Consider first the sentence in (5). Akan is a SVO language (Boadi, 1974; Saah, 1988; Ameka, 1992; Marfo & Bodomo, 2005; Kobele & Torrence, 2006). In (5) Kòfí is the subject, Á!uniG1C3má the object, and the verb is in progressive aspect as indicated by the L-toned prefix rè (cf., Dolphyne, 1988).

(5) Kòfí rè-bðá Á!má.

  Kofi PROG-help Ama

  ‘Kofi is helping Ama.’

(Marlo & Bodomo, 2005, p. 185; exclamation mark refers to lexical downstep)

From an information structural point of view, the sentence in (5) represents a neutral statement which we consider as being uttered in an out-of-the-blue context (cf., Ladd, 1980).
A sentence uttered without any context or as an answer to a very general question of the type “What happened?” is referred to as broad or wide focus (Ladd, 1980; Büring, 1997; Krifka, 2008). We will use the term wide focus here to refer to our neutral statements in the speech materials, which are elicited without context questions.

Contrary to a neutral statement, a focus is elicited by a context question asking explicitly for a particular constituent of the sentence. The element in question belongs to a set of alternatives (cf., Rooth, 1992, 1995; Krifka, 2008). This particular constituent receives pragmatic prominence and is syntactically focus-marked (Selkirk, 1995; Truckenbrodt, 1995). In Akan, *ex situ* focus is encoded syntactically by constituent fronting to the sentence initial position and morphologically by a focus marker (Boadi, 1974; Saah, 1988; Marfo & Bodomo, 2005; Ermisch, 2006; Kobele & Torrence, 2006; Amfo, 2010). In (6) the object Álmá is the answer to the preceding question, hence focused and syntactically fronted. In addition to focus fronting, the L tone focus marker nà obligatorily follows the fronted constituent. According to Boadi (1974) the focus marker nà has a semantic function of putting the constituent in contrast with a set of discourse alternatives, which in our view constitutes a focus in the sense of Rooth (1985, 1992). A further feature of focus fronting in Akan is pronoun resumption in case of animate focused elements (Boadi, 1974; Saah, 1988; Ermisch, 2006; Amfo, 2010). According to Kobele and Torrence (2006) this feature however is optional, and hence does not occur in our speech materials.

(6) a. Kófi ré-bóá nènà

Kofi PROG-help who

‘Who is helping Kofi?’
While all papers on focus in Akan agree that focus is expressed *ex situ* syntactically by means of focus fronting, Marfo and Bodomo (2005, p. 187) explicitly state that “a constituent cannot be contrastively focused *in situ* in Akan”. The authors argue that the focus marker *nà* represents the head of the extra-sentential focus phrase and does thus not appear *in situ*. However, focus may be realized *in situ* as is shown in (7) (Saah, 1988; Ermisch, 2006). According to Saah (1988) the answer in (7) presupposes the existence of an individual which has been asked for in the question, yet is less emphatic compared to an *ex situ* answer. Saah’s analysis is that *in situ* focus is a possible strategy (7).

(7)  a. Kòfí hù-ù hénà wò fiè hɔ?

Kofi see-PST who LOC house LOC

‘Whom did Kofi see in the house?’

b. Kòfí hù-ù [Kwàmé]₇ wò fiè hɔ

Kofi see-PST Kwame LOC house LOC

‘Kofi saw [Kwame]₇ in the house.’

(Saah, 1988, p. 25; tone and focus-marking added to the original)
Building on the work of Saah (1988), Ermisch (2006) differentiates between informational focus and identificational focus, which corresponds roughly to informational and corrective focus as we use the terms here (see below). In her view, these two focus categories are expressed differently in the syntax, the former with *in situ* focus, and the latter *ex situ*. The instances of *in situ* focus in Saah (1988), Kobele and Torrence (2006), and Ermisch (2006) do not show any occurrence of the focus marker *nà*. The objection put forward by Marfo and Bodomo (2005) was too strong. While preparing the materials for the present study our informants indeed did accept question-answer pairs with *in situ* focus answers.

To address the question of naturalness of the speech data of the present study and to check for any syntactic preference of prominence marking in Akan we conducted a small-scale situation-description task which was carried out before one of the production experiments (Genzel & Kügler, 2010). Pictures illustrating the situations of the mini-dialogues used in that production experiment (cf., (8) and (9) below) were shown to the speakers of this study and they were asked to answer simple wh-questions that elicit a focus structure. The speakers got no prepared answers as in the recordings of the experimental materials but were asked to answer spontaneously. With this task we would expect the speakers to choose their most frequent strategy to answer a question such as ‘What did Addo buy this morning?’ The test comprised two pictures referring to the two target words of the first production study for each of the eleven speakers. In the answers to an informational focus question no instance of an *ex situ* construction was realized as predicted by Ermisch.

---

6 The situation-description task was based on similar tasks proposed in the Questionnaire on Information Structure QUIS (Skopeteas et al., 2007). The task was adapted for the purposes of the present study (e.g., matching of target words).

7 One reviewer was asking for potential syntactic priming effects (cf., Bock, 1986). We do not see any influence of syntactic priming on the basis of this task and the corresponding answers. The wh-question asking for an informational focus had a syntactically fronted wh-word which was also followed by the focus particle *nà*. The corresponding question eliciting corrective focus had the contrasted element *in situ*. If syntactic priming would matter here, we would have expected more *ex situ* answers for informational focus, but the reverse was the case. In addition, we found an overall preference for *in situ* focus marking.
(2006), and in the answers to a corrective focus question three out of 22 answers were realized\textit{ex situ} (for details see Genzel & Kügler, 2010). These figures show that speakers frequently use the syntactically unmarked structure, which is in line with Saah (1988) and Ermisch (2006). Thus, the experimental set-up of the present study reflects a frequent use of prominence realisation. The fact that we find a similar prosodic effect for both \textit{in situ} and \textit{ex situ} focus constructions further supports \textit{in situ} focus occurrence as a natural way of expressing prominence in Akan.

To be more precise with the concept of focus we need to differentiate between different focus types. According to Krifka (2008) the category of focus can be divided into ‘informational’ and ‘corrective’ focus. The former focus type is shown in (7) with a wh-question asking for a particular constituent. Corrective focus on the other hand requires an antecedent in the previous discourse that the focus of the sentence would correct. A corresponding context for (7) would be the question “Did Kofi see Addo in the house?” For the present study we will test both informational and corrective focus. We assume an increase of prominence from informational to corrective focus, which corresponds to the discussion of different degrees of emphasis in Boadi (1974) and Saah (1988), or different focus types in Ermisch (2006). From languages such as German, we know that an increase of focus prominence is correlated with a gradual increase of prosodic cues such as pitch height and duration (Baumann, Grice & Steindamm, 2006).

Besides focus, other elements of a sentence are in the background (cf., Halliday, 1967; Krifka, 2008). We assume these elements to be given, defined as previously mentioned in the discourse. In the context question for informational focus (7) above both \textit{Kofi, in the house} and the verb are mentioned in the question, thus given. For the present study we will not only investigate the expression of prominence, in other words focus, but also any effect of

\footnote{Informational and corrective focus corresponds to what Ladd (1980) classifies as narrow focus. The concept of corrective focus often includes the notion of contrastive focus. For a detailed discussion about the concept of contrast see Zimmermann (2008), Repp (2010).}
givenness, hence no or reduced prominence. Prosodically, givenness may be accompanied by deaccentuation (cf., Cruttenden, 2006 for a range of languages, in particular Indo-European intonation languages) or pitch register compression (Xu, 1999 for Mandarin Chinese; Patil, Kentner, Gollrad, Kügler, Féry & Vasishth, 2008 for Hindi). Since no detailed phonetic data on Akan exist the present study intends to analyse the phonetic correlates F0 and duration in relation to information structure.

**Research question and hypotheses**

The goal of this study is to investigate the impact of information structure on the tonal realisation of L and H tones in Akan. This will provide insights into the prominence realisation in a tone language, and, more general, add to the question of the universality of the prosodic expression of prominence. We will consider both informational and corrective focus as well as givenness. Since Akan employs two syntactic constructions to express focus the comparison will include *in situ* and *ex situ* prominence marking. A neutral sentence as wide focus for each syntactic constructions matched for sentence length (syllables and tone) will serve as a baseline for comparison.

According to Boadi (1974) raising of H and L tones occurs in case of *ex situ* focus. The question remains whether this prosodic effect is also found in case of *in situ* focus. If it is the case that tonal raising is the prosodic correlate of focus in Akan we expect it to be realized *in situ* as well.

The impact of givenness on the prosodic structure is unknown. Two distinct effects may arise: (i) no change in tonal realisation, (ii) compression of pitch register as in Mandarin Chinese (Xu, 1999) or Hindi (Patil et al., 2008). Deaccentuation as in intonation languages (cf., Cruttenden 2006) is not expected since lexical tone distinctions have to be maintained in a tone language.
On the basis of Boadi (1974) we assume that Akan belongs to the group of languages that uses prosodic means for the expression of focus. We therefore predict that Akan is comparable to languages such as German, English or Mandarin Chinese. From a typological point of view the analysis of a terraced level tone language like Akan thus adds data to the theory of intonational universals.

PRODUCTION EXPERIMENTS

For the present study data from three different production experiments is presented. The speech materials as well as the speakers differ across the experiments; however the elicitation of focus on particular elements of the experimental sentences by means of questions was identical throughout the three experiments. Based on the unexpected findings of experiment 1, that is the pitch register lowering under corrective focus, we conducted two further production studies with different speech materials and different speakers to provide more data that corroborates the analysis.

Method

Design and Materials. The first experiment (experiment 1) involved three factors: information structure (wide focus, informational focus, corrective focus, pre-focal and post-focal givenness), syntactic construction (in situ or canonical word order as in (7), and ex situ as in (6) above), and tone (L, H). Two target words, one carrying lexically L tones (Âddò, (8)) and one carrying a H tone (àmáŋgò, (9)), were embedded in carrier sentences.

(8) Àgyèmàñ bóá-á Âddò ánðpá yí.
Agyeman help-PAST Addo morning this
‘Agyeman helped Addo this morning.’
(9) Ànúm̀ tò-à àmáŋgò ánòpá yí.
Anum buy-PAST mango morning this
‘Anum bought a mango this morning.’

To elicit the desired information structure of a sentence we used mini dialogues. Each sentence was preceded by a question except for sentences in wide focus. The question-answer task allows for a direct control of the presupposition of the proper target sentence which in turn determines the focus structure of a sentence. Wide focus was elicited without a context and the whole sentence is assumed to be in focus. Examples (8) and (9) show the sentences in the wide focus condition with an *in situ* target word.

The questions set up either an informational (10) or a corrective focus (11) on the target word for both syntactic constructions, or the target word in pre-focal (12) or post-focal (13) position for *in situ* occurrences. The examples (10) – (13) illustrate a question–answer set for the *in situ* target word carrying a L tone. The same focus structures are applied for sentences containing the target word carrying the H tone (9).

(10) a. Hwàń nà Àgyèmà bóà-à ánòpá yí?
   whom FM Agyeman help-PAST morning this
   ‘Whom did Agyeman help this morning?’

b. Àgyèmà bóà-à [Àddò] bóà-à ánòpá yí.
   Agyeman help-PAST Addo morning this
   ‘Agyeman helped Addo this morning.’

(11) a. Àgyèmà bóà-à Ànúm̀ ánòpá yí.
   Agyeman help-PAST Anum morning this
The second syntactic construction consists of sentences with the two target words in sentence initial position, with the target word fronted due to focus. In (14) and (16) the target words carrying L and H tone appear in focus position, with the obligatory morphological focus marker nà. To make this kind of prominence explicit the literal translation in English contains a cleft construction. The original Akan sentence is not a syntactic cleft construction. In order to identify any prosodic effect of focus in this particular position the phonetic properties of the target word needs to be compared with sentence initial non-focused target words. These are given in (15) and (17) for the target word carrying L and H tone respectively. The target words appear as canonical pre-verbal subjects.

The whole data set of experiment 1 consisted of 16 sentences (5 information structures x 2 target words in situ, plus 3 information structures x 2 target words ex situ).
The second experiment (experiment 2) replicated the first experiment with a different group of speakers concentrating on the *in situ* focus constructions. Two factors (information structure and tone) were studied to gain more evidence for the unexpected focal lowering observed in experiment 1. As for information structure the sentences were uttered without a context question eliciting wide focus, and with context questions eliciting both informational and corrective focus. One disyllabic word carrying lexical H tones (18), and one trisyllabic word carrying lexical L tones (19) on each syllable were embedded in a carrier sentence that was controlled for lexical tone and that resembles the one of experiment 1.
(18) Afùà hùnùù wómá ánòpá yí.
Afua see.PAST pestle morning this
‘Afua saw a pestle this morning.’

(19) Afùà hùnùù àtèrè ánòpá yí.
Afua see.PAST spoon morning this
‘Afua saw a spoon this morning.’

A third experiment (experiment 3), again with a different group of speakers, was designed to test for the focal lowering in a different sentence frame in combination with downstep. The target word, that is the proper name in (20) has the tonal specification L-H-L-H, which presents the environment for downstep. Remaining tones in the sentence are low in order to not intervene with other H tones, except for the last H tone which phonetically turns into a L tone at the end of an utterance (Dolphyne, 1988; cf., also Good, 2004 for a similar phenomenon in Saramaccan). The sentence in (20) was recorded without context, and with a context question asking “Is it Kofi Annan speaking?” which contrasts the surnames ‘Annan’ and ‘Gyima’.

(20) è-yè Kòfí Gyìmá nà 3-rè-kàsá
PRO-be.PRES Kofi Gyima FM PRO-PROG-speak
‘It is Kofi Gyima speaking.’

Participants. In all three experiments speakers were native speakers of Asante Twi as spoken in and around Kumasi. All speakers declared English as their second language. Eleven speakers (6 female and 5 male) participated in experiment 1. Eight participants were students
of the University of Ghana in Accra. As for the other three, one was in the army and based in Kumasi, another worked in a pharmacy, and the third one worked as a university lecturer. The average age was 26 years.

In experiment 2, five speakers (4 male, 1 female) who study in Minot, North Dakota, participated. Their average age was 28. Five speakers (3 male and 2 female) participated in experiment 3. Their average age was 32 years. All of the speakers lived in Berlin working in different companies or were exchange students at a university in Berlin. Each speaker was paid a small fee for participation.

*Recording procedure.* The experiments were carried out using presentation software. The question-answer trials were prepared in Akan orthography. The questions spoken by a young female were pre-recoded in a quiet room in Berlin directly onto a laptop using Audacity (Version 1.2.6) and a headset (Logitech Internet Chat Headset). The question-answer pairs were presented to each speaker in a pseudo-randomized order. Items from two other experiments, unrelated to these experiments, were interspersed as fillers.

Participants were digitally recorded on a laptop via a headset (Logitech Internet Chat Headset) using Audacity (Version 1.2.6). The headphones were binaural with a frequency spectrum from 20-20000 Hz and an acoustic impedance of 32 Ohm with an integrated volume control, so that every participant could adjust the volume. The microphone was an electret condenser type with a sensitivity of -39 dBV/Pascal. The first experiment was carried out in a quiet room in the Linguistics Department at the University of Ghana, the second in a quiet room in Minot, North Dakota, and the third in a quiet room in Berlin.

The participants were familiarized with the task through written and oral instructions, followed by four practice trials. Trials consisted of a visual presentation of the question and its answer on a computer screen, or only the statement in the case of wide focus. In each wide focus condition only the target sentence was presented. Participants heard the pre-recorded
question over headphones. Simultaneously the target sentence was presented on the screen. The participants were instructed to first listen to the question and to subsequently read the answer quietly. Then, the question was presented again and the participants had to read the answer out loud as a response to the question. If the answer started with *dáàbí* ‘no’ the participant was asked to put a pause after it. If the question was answered without any hesitations, false starts or missing phrasing, the next trial was presented. If there were hesitations the participants were asked to repeat the answer which was the case for 28 out of 264 items (11%) in the first experiment and about equivalent ratios in the other two experiments. Presentation flow was controlled by the experimenter and participants were allowed to take a break whenever they wanted. In total each of the experiments lasted approximately 45 minutes.

*Data pre-processing and statistical analysis.* The recordings were digitized at a sampling frequency of 44.1 kHz and 32 bit resolution. The target word in all sentences was labelled by hand at the level of the word and the syllable. Labelling in Praat (Boersma & Weenink, 2008) was based on a see-listen-label method, in other words visually evaluating the spectrogram and listening to the sound files. Standard cues for segmental labelling were employed (Turk, Nakai, & Sugahara, 2006) and boundaries were set automatically at zero crossings using a Praat script. For each target word the durations of the word and the syllables were extracted using a Praat script.

The F0 analysis was based on a Hanning window of 0.4 seconds length with a default 10 milliseconds analysis frame. Every pitch object was visually checked for octave jumps and algorithm faults and manually corrected. Applying the Praat smoothing function the F0 was smoothed at 10Hz to diminish microprosodic perturbations. Two different measurements were conducted. First, for each syllable of the target word the time-normalized course of the F0 contour was extracted by measuring the F0 at ten equal steps during each syllable. Resulting
F0 measures were aggregated per speaker and condition. The aggregated scores were plotted over time steps interpolating between the individual measures. This type of measurement provides phonetic details of the course of the F0 contour. Second, tones were labelled for every target word. The tonal marker was set manually at the middle of each of the tone bearing vowels. We assume the mid-point of each TBU to reflect a phonetic interpretation of a discrete lexical tone. The corresponding F0 was extracted in Hertz (Hz).

The statistical analysis relied on the dependent variables ‘F0 (Hz)’ and ‘duration (ms)’. The scores were subjected to repeated measures ANOVAs with speaker as random factor. Further, the obtained F0 values were aggregated within each participant and each condition to calculate the difference (Δc) of F0 and duration between the wide focus baseline realisation and any of the four information structure conditions for the first experiment. For the other two experiments we calculated the difference between the wide focus baseline and the corrective focus conditions.

**RESULTS**

The analysis of information structure effects on the tonal and durational realisation comprises two different dimensions of prominence. First, an increase of prominence is expressed from informational to corrective focus compared to neutral wide focus sentences. Second, a decrease in prominence is expressed if the target word is given and the focus of the sentence appears either before or after the target word. In the former case, the target word occurs as post-focally given, in the latter as pre-focally given. The presentation of the results is organized by experimental factors. We will first report the effect of *in situ* focus on the tonal and durational realisation across the three experiments. Thereafter we will present results of the effect of syntactic focus construction on the tonal and durational realisation comparing *in situ* and *ex situ* focused target words of experiment 1. Finally we will elaborate on the effect
of givenness on the tonal and durational realisation of experiment 1. The presentation will consider H and L tones separately.

**The effect of in situ realized focus on tone and duration**

Overall, a gradual decrease in F0 height can be observed with increasing prosodic prominence. Figure 2 shows the course of F0 for the target word àmáŋgõ of experiment 1. The second syllable carries a lexical H tone. The solid line displays the wide focus condition which serves as the baseline for comparison. The dotted line represents the target word in informational focus, the dashed line in corrective focus. In all three conditions F0 increases towards and decreases from an F0 maximum on the syllable that carries the lexical H tone.

### Insert Figure 2 about here

The scaling of the H tone differs as a function of information structure. It is realized lower if focused. The aggregated mean F0 for wide focus is at 181 Hz, for informational focus at 172 Hz, and for corrective focus at 166 Hz (cf., Table 1). On average the lowering of corrective focus *in situ* amounts to 1.5 st. Equivalent figures are found in the other two experiments (cf., Table 1), though the lowering effect only amounts to 1.2 st in experiment 2 and to 0.7 st in experiment 3.

### Insert Table 1 about here

To test for a significant effect of focus in experiment 1 a 3x2 repeated measures ANOVA was performed with Focus and Syntactic Construction⁹ as fixed factors and Subject as random factor. The analysis revealed a significant main effect of Focus, $F(2, 10) = 17.4$, $p$
< 0.001 for the H tone. Post-hoc t-tests showed no significant effect for informational focus compared to the wide focus baseline, but a significant lowering of corrective focus compared to wide focus for *in situ* focus, *t*(1, 10) = 3.0, *p* < 0.05. A repeated measures ANOVA for experiment 2 with Focus as the fixed factor and Subject as the random factor revealed a significant lowering of the H tone in case of corrective focus compared to wide focus, *F*(1, 4) = 10.0, *p* < 0.05. An identical repeated measures ANOVA for experiment 3 with Focus as the fixed factor and Subject as the random factor revealed also a significant lowering of the H tone in case of corrective focus compared to wide focus, *F*(1, 4) = 14.6, *p* < 0.05. Recall that in experiments 2 and 3 the factor Focus comprises only of two levels, that is the distinction between wide and corrective focus.

The durational figures showed no consistent pattern across experiments. The duration of the target word carrying the H tone in experiment 1 was on average 58 ms shorter in the corrective focus condition compared to the wide focus. The durational reduction amounts to 10.3 % of the average word duration in wide focus. A 3x2 repeated measures ANOVA that was identical to the one carried out for F0, was carried out for duration. The factor Focus showed a main effect, *F*(2, 10) = 9.1, *p* < 0.01. Post-hoc t-tests only showed a significant reduction of duration in case of corrective focus compared to wide focus, *t*(1, 10) = 2.5, *p* < 0.05. However for experiments 2 and 3 no significant difference in duration between focus conditions was observed.

Figure 3 shows the course of F0 on the target word *Addò*. Again, the solid line displays the wide focus condition which serves as the baseline for comparison. For the target word carrying L tones a similar gradual decrease in F0 height with increasing prosodic prominence can be observed as for the target word carrying a H tone. On the first syllable F0 falls towards a L tone target. F0 on the second syllable remains rather flat at a level of, on average, between 130 to 140 Hz. Generally it becomes clear that the second lexical tone is
realized lower than the first one. This seems mainly be due to the fact that the F0 on the first syllable is a transition from a preceding High tone to the low tone target associated with this syllable (cf., (8)).

### Insert Figure 3 about here

Again, a 3x2 repeated measures ANOVA was performed with Focus and Syntactic Construction as the fixed factors and Subject as the random factor. The factor Focus approached significance, $F(2, 9) = 2.9, p = 0.077$. Post-hoc t-tests only showed a significant lowering of corrective focus compared to wide focus for in situ focus, $t(1, 9) = 2.9, p < 0.01$. On average the lowering was about 1.0 semitone (cf., Table 1).

Given the lowering effect of L tones in corrective focus compared to the wide focus baseline the question remains whether both L tones equally participate in the lowering effect. To test this, a further repeated measures ANOVA was performed with Syllable (first vs. second syllable carrying a lexical L tone) as the fixed factor and Subject as the random factor. The analysis revealed a significant difference between the two lexical L tones, $F(1, 9) = 7.3, p < 0.05$. Post-hoc t-tests showed then that only the second of the two L tones was significantly lower in the corrective focus condition compared to the wide focus condition, $t(1, 9) = 4.2, p < 0.01$.

In experiments 2 and 3 corrective focus causes a similar lowering effect on L tones, yet the comparisons did not yield any significance, $F(1, 4) = 3.1, p > 0.05$ for experiment 2, $F(1, 4) = 1.2, p > 0.05$ for experiment 3. In experiment 3, the L tone precedes the downstepped H tone, and may therefore not be systematically affected.

---

10 One of the eleven speakers had creaky voice during the second syllable thus no F0 measurement was obtained.
Given the different participation of L tones in the lowering found in experiment 1 we also performed a repeated measures ANOVA with Syllable (first vs. second vs. third lexical L tone) as the fixed factor and Subject as the random factor for experiment 2 (target word àtèrè ‘spoon’) comparing wide focus with corrective focus. If only the last L tone may participate in the lowering effect as in experiment 1 the insignificant main effect for Syllable may be masked by the fact that the target word is realized with three L tones. Thus, the ANOVA tests any participation of the individual lexical tones in the lowering effect. The analysis revealed a highly significant effect of Syllable, $F(2, 4) = 35.8$, $p < 0.001$. A repeated measures ANOVA for each of the three tones with Focus as the fixed factor, and Subject as the random factor only showed a significant lowering of the third lexical L tone, $F(1, 4) = 8.6$, $p < 0.05$.

As far as duration is concerned in none of the three experiments we obtained a significant effect for the target word carrying L tones, and no consistent pattern was observed. The durational change between the wide focus baseline and the corrective focus condition in experiment 1 was a reduction of 15 ms on average, while it was an increase of 20 ms on average in experiment 2, and no difference in experiment 3.

**The effect of ex situ realized focus on tone and duration in comparison to in situ focus**

The analysis so far considered the occurrence of target words in sentence medial, in particular post-verbal position which corresponds syntactically to the canonical or neutral word order in Akan (e.g., Kobele & Torrence, 2006). For the expression of prominence speakers of Akan may however choose a different syntactic construction which is accompanied by morphological focus marking (cf., (6b)). To compare any effect of ex situ focus two wide focus sentences with the target words appearing in sentence initial position served as a baseline for comparison in experiment 1.
Figure 4 shows the time-normalized course of F0 aggregated for all speakers for the target word carrying the H tone appearing in situ (black lines) and ex situ (grey lines). The solid lines correspond to the in situ and ex situ wide focus, respectively. Dotted lines correspond to informational focus, and dashed lines to corrective focus conditions. Two facts become obvious from that comparison. First, the sentence initial or ex situ rendition of the target word is overall higher in F0 than the in situ one. Syllables carrying a L tone are on average 30 Hz higher, the syllable carrying a H tone is on average 80 Hz higher. The 3x2 repeated measures ANOVA with Focus and Syntactic Construction as fixed factors and Subject as random factor exposed a significant main effect of Syntactic Construction, $F(1, 10) = 41.0, p < 0.001$. The fact that the overall higher realisation of ex situ target words compared to in situ ones is significant is expected in a terraced-level tone language since speakers start in a higher pitch register which decreases throughout the sentence (cf., Figure 1 above).

Second, the H tone in the corrective ex situ focus rendition is lower than in informational focus and in the wide focus condition. Post-hoc t-tests show that the comparison between the baseline and the informational focus condition yields no significance, $t(1, 10) = 0.7, p > 0.05$ while the average lowering of 1.8 st (cf., Table 1) in corrective focus compared to the wide focus baseline is significant, $t(1, 10) = 5.5, p < 0.001$. The amount of lowering in the ex situ construction (1.8 st) is higher than in the in situ construction (1.5 st) (cf., Table 1).

The durational patterns in sentence initial position differ from those of the in situ construction. Overall, no shortening of word duration between the wide focus baseline and the two focus conditions is observed. On average the durational differences are very small, 8 ms.
lengthening in informational focus and 7 ms shortening in corrective focus. The 3x2 repeated measures ANOVA did not reveal any significant differences for duration.

Also for the target word carrying L tones the overall higher realisation of *ex situ* target words compared to *in situ* realisations can be observed. The average difference between these two is about 30 Hz across conditions. The 3x2 repeated measures ANOVA with Focus and Syntactic Construction as the fixed factors and Subject as random factor showed a significant main effect of Syntactic Construction, $F(1, 9) = 87.1, p < 0.001$.

Comparing the L tone in *ex situ* position across focus conditions we observe an overall lowering (cf., Table 1). As reported above, the factor Focus in the 3x2 repeated measures ANOVA with Focus and Syntactic Construction as fixed factors and Subject as random factor approached significance, $F(2, 9) = 2.9, p = 0.077$. Post-hoc t-tests for *ex situ* focused words however did not reveal any significant difference between the wide focus condition and informational focus, $t(1, 9) = 0.4, p > 0.05$, or between wide focus and corrective focus, $t(1, 9) = 1.8, p > 0.05$.

The durational patterns of the L tone target word in sentence initial position did not show any significant difference. On average the target word shortened about 5 ms in informational focus to and 16 ms in corrective focus compared to the wide focus condition.

**The effect of givenness on tone and duration**

If the preceding context question contains a word or constituent that is repeated in the target answer sentence, this element is considered to be given. At the same time the focus of the sentence appears elsewhere in that sentence. Either the focus appears before the target word or after it. Figure 5 shows the course of F0 on the target word àmáŋgò comparing the wide focus condition as the baseline with the target word when pre-focally given, and when post-focally given. We observe a difference between a pre-focal and post-focal target word. The syllable
of the target word carrying the H tone in the post-focal condition is realized lower than both in the wide focus and pre-focal condition (Figure 5).

A repeated measures ANOVA for the H tone with Givenness as the fixed factor and Subject as the random factor revealed a significant main effect of Givenness, \( F(2, 10) = 9.2, p < 0.001 \). Post-hoc t-tests exposed that only the post-focal givenness condition showed a significant lowering effect of 1.4 semitones on average compared to wide focus, \( t(1, 10) = 4.1, p < 0.005 \). In pre-focal position there was no significant lowering effect compared to wide focus.

The overall word duration decreases with a decrease in prominence assuming that given words are less prominent. Comparing with the mean word duration of 536 ms in wide focus, pre-focally given words are on average 21 ms shorter in duration, while the shortening of post-focally given words amounts to 51 ms on average. However, a repeated measures ANOVA with Givenness as the fixed factor and Subject as the random factor revealed no significant main effect.

For the target word carrying L tones, ‘Addö’, no significant effect of givenness was observed, \( F(1, 9) = 1.4, p > 0.05 \). Also for duration, the repeated measures ANOVA yielded no significance, although the means across speakers show a durational decrease of 17 ms on average in post-focal condition, and 29 ms in pre-focal condition.

**DISCUSSION**

**Intonation in tone languages**

This study presented the results of three production experiments on Akan, a two-tone language spoken in Ghana. The main question under investigation was whether sentence-level
pragmatic meanings (Ladd, 1996) such as focus and givenness may be prosodically encoded in a tone language despite the fact that the phonetic correlate F0 is functionally used to distinguish lexical and grammatical meanings of tones. Languages indeed differ with respect to the prosodic properties they use for the expression prominence or focus if at all. In the introduction we discussed that languages do differ as a function of their typological profile in the use of prosodic properties. For intonation languages like Wolof, for instance, no effects of focus on the intonation have been found (Rialland & Robert, 2001) while for other intonation languages such as German and English numerous studies exist that show such effects. On the other hand certain tone languages, for example Mandarin Chinese (Xu, 1999, 2005), show similar effects as intonation languages in the prosodic expression of focus.

The expectations for Akan were that speakers employ prosodic means for the expression informational and corrective focus (Boadi, 1974). The results of the present study indeed corroborate the expectation, yet in an unexpected way. While the hypothesis based on Boadi’s (1974) claim was that the prosodic expression of focus would result in tonal raising as in Mandarin Chinese, English or German, the data of the three independent production studies showed a pitch register lowering effect in case of corrective focus. This effect is robust for lexical H tones, and only marginally observed for L tones. In the following, this effect will be discussed in light of the theory of the biological codes of intonation.

**Deviation from a neutral register**

The main finding of this paper is that the prosodic expression of prominence goes against expectations and, more striking, against the prosodic effect of tonal or register raising and durational lengthening in relation to focus (e.g., Ladd, 1996; Gussenhoven, 2004; Féry & Kügler, 2008). In Akan, we find a gradual decrease in F0 height with increasing prosodic prominence. However, duration was not affected in a systematic way. Comparing a pragmatically neutral rendition of a target word embedded in a sentence with renditions of the
same target word in informational and in corrective focus we observed pitch register lowering for lexically H and certain L tones. The lowering was significant only in case of corrective focus, but not significant in case of informational focus.

In light of a theory of the biological foundations of intonation such as the biological codes (Gussenhoven, 2002, 2004) the results of Akan pitch register lowering appear to go against an assumption made based on the effort code. According to the effort code putting more effort into the speech production mechanism correlates with more precise supralaryngal articulatory movements (de Jong, 1995) and less undershoot of articulatory targets in terms Lindblom’s (1990) hyper- and hypo-speech theory of speech production. The more precise articulation does also affect the laryngeal coordination and thus the properties of tone and intonation which in turn means an expansion of pitch movements, hence the pitch register (Gussenhoven, 2002, 2004). Gussenhoven relates this effect with emphasis, a general communicative goal of a speaker to get his message across. From a grammatical point of view the effort code is related to the expression of focus which is phonetically realized by means of relatively wide F0 excursions. In particular the phonetic and phonological details of the prosodic expression of focus revealed a tonal or register raising in Germanic languages such as English (Cooper et al., 1985; Eady & Cooper 1986; Eady et al., 1986), Dutch (Swerts, Krahmer, & Avesani, 2002), and German (Baumann et al., 2006; Baumann, Becker, Grice & Mücke, 2007; Féry & Kügler, 2008).

The prosodic expression of focal prominence by means of pitch register lowering contradicts the idea of the effort code which predicts its deviation from a neutral voice only in one direction, in other words an expansion of pitch register which causes a raising of tones. On the basis of the present findings we would suggest an alternative view on the prosodic expression of focus. Given the fact that the prosodic expression of focus differs from a neutral register (wide focus) it seems plausible that it is not the direction of change, in other words...
raising as in Germanic languages and predicted by the effort code, that does matter but the deviation from the neutral register.

Since the durational measurements did not show a consistent pattern across the three experiments the observed durational reduction of targets words of experiment 1 does not seem to correlate with the effect of pitch register lowering. However, one may assume a correlation between segmental duration and F0. If focus causes a durational reduction of segments there is less time to reach a specific tonal target. According to Xu and Sun (2001) the vocal folds need a certain amount of time to accommodate a higher or lower F0. Less time or shorter duration means thus that any continuous F0 change results in lower F0 maxima or higher F0 minima. Since we did not find a consistent durational reduction in our data we assume no direct correlation between the segmental duration and the realisation of lower H and L tones under focus. Also the inconclusive durational patterns do not hint at target undershoot as would have been a prediction of Lindblom’s (1990) hyper- and hypo-speech theory of speech production. The precision in articulation still remains, but the pitch register is lowered, which causes the tonal target to be realized lower. For the target words carrying L tones the lowering effect would also go against the undershoot view since this would predict somewhat higher L tones. The duration of the target word thus does not constrain the F0 height of the lexical tone. This fact points to the fact that the deviation of F0 in corrective focus cannot be a consequence of a durational change.

A further hint for the argument that the deviation from a neutral pitch register matters regardless of which direction the deviation goes, comes from languages that use prosodic phrasing as a prosodic cue for the expression of focus, for instance Chichewa (e.g., Kanerva, 1990; Downing et al., 2004). According to Downing et al. (2004, p. 184) focused words or phrases in Chichewa are recognized because they deviate in some sense from an unmarked prosodic structure. This deviation from a neutral phrasing is the crucial point of view in relation to an effort driven theory.
Evidence from unrelated languages also shows that corrective focus is realized in strikingly different ways (Ladd, 1996, p. 177). According to Ladd, Italian, for instance, allows for deaccentuation of larger constituents although deaccentuation of post-focal constituents usually does not apply (Swerts et al., 2002). Moreover, Akan does not seem to be an exceptional case employing pitch register lowering in the prosodic expression of corrective focus. Gili-Fivela (2008) reported a similar lowering for Italian in correlation with corrective focus. Even for English Cooper & Sorensen (1981, p. 3) report on that focus occasionally is expressed with “unusual low F0” (cf., Liberman & Pierrehumbert, 1984, for a similar prediction). And also the data by Xu (1999, 2005) on Mandarin Chinese does not only reveal tonal raising, but also lowering for the L tone. Thus, the fact of lower pitch register as a prosodic expression of focus cannot be regarded as an exceptional case of language variation. And given the idea that a deviation from neutral register may guide listeners’ attention to central informational parts of an utterance, more languages are expected to employ deviant patterns.

Anecdotal evidence for the deviation account comes from classroom experience at the university level. As a teacher, one may raise one’s voice to attract attention of students mumbling or whispering. However, the first author frequently employs the opposite strategy of lowering his voice to achieve the goal of directing the attention towards the content of the lecture. This example simply shows that the communicative function of highlighting information that is correlated with intonation can be achieved by means of different strategies.

On the universality of the prosodic expression of givenness
This study not only tested the prosodic expression of focus but also any effect of givenness on the tonal realisation. The property of givenness arises through a focus elsewhere in the sentence. Languages appear to distinguish between these different positions with respect to givenness. Cruttenden (2006) surveyed a great variety of languages that do show post-focal
deaccenting, which means that no tonal correlates after the focus appear. Yet, languages like Hindi or Mandarin Chinese, for instance, do not show a complete deletion of tonal cues after a focus but post-focal register compression that still allows tones to be realized (Patil et al., 2008; Xu, 1999). Since tone languages rely on the expression of tonal distinctions for lexical and grammatical reasons we would not expect complete tonal deletions after a focus. However, register changes as is the case in Mandarin Chinese may well be an issue. In contrast to post-focal position, pre-focally given words, constituents or phrases may or may not be prosodically marked. In German, for instance, pitch accents may well be realized before a focus (Baumann, 2006). If pre-focally realized the scaling of these accents appears in a lower pitch register, an effect which Féry and Kügler (2008) relate to givenness.

In Akan we do not find register effects exclusively attributable to givenness. The position of the focus towards the given element is relevant for its prosodic realisation since a distinction between pre-focal and post-focal givenness with respect to its prosodic expression is observed. For lexical L and H tones associated with words occurring in pre-focal position we did not find any significant difference in tonal realisation compared to the neutral realisation. This observation rules out an alternative explanation for the focus lowering effect. According to one reviewer an alternative interpretation of the pitch register lowering due to focus reported here could be a global register lowering affecting the whole utterance. The global lowering may arise if one assumes an influence from the negation marker dààbí ‘no’.

During the data elicitation, speakers were asked to clearly pause after the negation marker. This instruction was given to ensure no further effect on following target sentence. All speakers acted according to this instruction, and evaluation of the recordings revealed a clear pause in all recordings.

The fact that pre-focal givenness does not show any significant lowering provides evidence against a global interpretation of register lowering. If the sentences containing a corrective focus would be consistently realised in an overall lower register as a consequence
of the negation marker, we would have been expected a clear pre-focal effect in our materials of experiment 1.

Figure 6 provides further evidence that the pitch register lowering effect is a local one rather than a global one. As an example from the second experiment, in Figure 6 the averaged course of F0 across all speakers of the whole utterance containing the target word carrying H tones in experiment 2 is shown. This sentence perspective on the F0 course as employed for instance by Remijsen and van Heuven (2005) allows for a conclusive decision of global or local implementation of the lowering effect.

### Insert Figure 6 about here

From the comparison of the wide focus renditions with that of corrective focus follows that the lowering takes place on the target word itself. The F0 of the previous H tone associated with the subject of the utterance is scaled exactly at the same height. In summary, the data of the present study suggests that the corrective focus induced by pitch register lowering is a local effect affecting the focussed word and the stretch of utterance after it.

Therefore we found a lowering of the pitch register in post-focal position for both lexically L and H tones. The observed post-focal lowering of tones compared to the wide focus baseline turns out to be a tonal realisation in a lower register that previously has been triggered by focal lowering. Thus, post-focal tones in Akan are lowered not because of the fact that givenness reduces any articulatory effort or any degree of prominence but because of focal lowering.

**Pitch register lowering and categories of focus in Akan**

Following Krifka (2008) we distinguish between informational and corrective focus throughout the paper. The latter has been shown to be prosodically realized by means of pitch
register lowering while the former does not differ from wide focus or out-of-the-blue utterances. From this distinction the question arises whether there is a concept such as informational focus in Akan.\footnote{Thanks to Bert Remijsen who pointed us to this issue.} Hartmann (2008) claims that there is only one category of focus and she presents evidence from German and Hausa supporting this claim. The variance between different types of focus then is claimed to be gradient. While this is certainly true for German, Akan clearly distinguishes between a corrective focus and other types of focus. One interpretation that might explain this sharp distinction could be the very fact that Akan as a tone language uses F0 primarily as a cue to express tone. Hence, F0 is not used to distinguish between a wide and an informational focus. In case of corrective focus, on the other hand, an additional pragmatic prominence comes into play which speakers may want to express even prosodically. This additional prominence is correlated with a stronger communicative goal to emphasize on a certain part of an utterance, and speakers of Akan draw attention to that kind of information by means of pitch register lowering.

**Positional effects on prosody**

A third conclusion of this paper concerns the comparison between \textit{in situ} and \textit{ex situ} focused elements. Our data does not support Boadi’s (1974) claim that both L and H tones are raised in \textit{ex situ} focus. On the contrary, Figure 4 clearly shows a significant F0-lowering in case of corrective focus both for \textit{in situ} and \textit{ex situ} cases. Thus, the prosodic expression of focal prominence by means of pitch register lowering is similar in different syntactic focus constructions and independent of any other linguistic marking of focus be it syntax and morphology in case of \textit{ex situ} focus.

Possibly, Boadi interpreted the overall higher pitch register at the beginning of a sentence as an effect of focus on the tonal realisation. In his data, Boadi did not compare the pretended raised tone with a baseline. Being overall higher at the beginning of an utterance
simply reflects the start of an utterance but does not express prominence. The tendency of speakers across languages to start higher at the beginning of an intonation phrase led Gussenhoven (2002, 2004) to propose the production code, which he assumes to be a physiological correlation of more subglottal pressure at the beginning of an intonation phrase. The data in our study very clearly disentangles the positional effect from the information structural effect in that we compare the corrective focus case in both focus constructions with an equivalent neutral baseline realisation.

**Summary**

Given the fact that Akan uses intonation, for instance to distinguish between questions and statements, the question of the present study was to what extent pragmatic prominence influences the scaling of tones, which is determined by the pitch register. We showed that pragmatic prominence induces pitch register lowering in the case of corrective focus independent of syntactic focus construction. The lowered pitch register is maintained in a phrase so that post-focal tones are scaled relative to the lowered register. The pragmatic meaning of contrast is expressed prosodically but the prosodic expression is constrained by the tonal phonology of Akan.
REFERENCES


Kügler, F., & Genzel, S. (submitted). Information structure, sentence length, and position effects on duration in German.


### Tables

<table>
<thead>
<tr>
<th></th>
<th>H tone</th>
<th>L tone</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>wFoc</td>
<td>cFoc</td>
<td>Δ (st)</td>
<td>wFoc</td>
<td>cFoc</td>
<td>Δ (st)</td>
</tr>
<tr>
<td>Exp 1</td>
<td>214 (68)</td>
<td>195 (70)</td>
<td>1.6 ***</td>
<td>153 (40)</td>
<td>145 (38)</td>
<td>0.9 n.s.</td>
</tr>
<tr>
<td>Exp 1 <em>in situ</em></td>
<td>181 (44)</td>
<td>166 (41)</td>
<td>1.5 *</td>
<td>143 (39)</td>
<td>135 (36)</td>
<td>1.0 **</td>
</tr>
<tr>
<td>Exp 1 <em>ex situ</em></td>
<td>248 (72)</td>
<td>224 (67)</td>
<td>1.8 ***</td>
<td>162 (40)</td>
<td>155 (37)</td>
<td>0.8 n.s.</td>
</tr>
<tr>
<td>Exp 2</td>
<td>150 (35)</td>
<td>140 (32)</td>
<td>1.2 *</td>
<td>134 (34)</td>
<td>127 (31)</td>
<td>0.9 n.s.</td>
</tr>
<tr>
<td>Exp 3</td>
<td>146 (32)</td>
<td>140 (30)</td>
<td>0.7 *</td>
<td>132 (29)</td>
<td>129 (28)</td>
<td>0.4 n.s.</td>
</tr>
</tbody>
</table>

Table 1. Mean F0 values in Hertz for H and L tones, averaged across speakers and tones with standard deviation in parenthesis, are given for the three different production experiments. The amount of F0 difference in semitones (Δ (st)) is calculated between the wide (wFoc) and corrective focus condition (cFoc). Significance levels refer to * < 0.05, ** < 0.01, *** < 0.001, n.s. not significant. The averaged data from experiment 1 is also split by syntactic focus construction, *in situ* and *ex situ* focus.
Figure titles

Figure 1: Tonal terracing at three distinct register levels exemplified with sentence (4). L tones between H tones lower the register in a staircase fashion illustrated by horizontal lines.

Figure 2: Time-normalized course of F0 by syllable on the target word carrying a H tone, àmángò, occurring in in situ focus position, aggregated for all speakers (n=11). The solid line corresponds to wide focus condition, the dotted line to informational focus, and the dashed line to corrective focus.

Figure 3: Time-normalized course of F0 by syllable on the target word carrying L tones, Àddò, in in situ position, aggregated for all speakers (n=11). The solid line corresponds to wide focus condition, the dotted line to informational focus, and the dashed line to corrective focus.

Figure 4: Time-normalized course of F0 by syllable on the target word carrying a H tone, àmángò, aggregated for all speakers (n=11). Black lines refer to the ex situ target word occurrence, grey lines to the in situ ones. The solid lines correspond to the baseline condition (wide focus), the dotted lines to informational focus, and the dashed lines to corrective focus.

Figure 5: Time-normalized course of F0 by syllable on the target word carrying a H tone, àmángò, in in situ position, aggregated for all speakers (n=11). The solid line corresponds to wide focus condition, the dotted line to pre-focal givenness, and the dashed line to post-focal givenness.

Figure 6: Time-normalized course of F0 by syllable for the sentence Àfiú àhùnù wó má ànòpá yì, aggregated for all speakers (n=5). The target word wó má carries two H tones. The target word is highlighted. The solid line corresponds to wide focus condition, the dashed line to corrective focus.
Figure 1
Figure 2
Figure 4
Figure 5
PHONETIC REALIZATION OF AUTOMATIC (DOWNDRIFT) AND NON-AUTOMATIC DOWNSTEP IN AKAN

Susanne Genzel & Frank Kügler

Department of Linguistics & SFB 632 “Information structure”, University of Potsdam
susonne7@gmail.com, kuegler@uni-potsdam.de

ABSTRACT
This paper deals with the question whether automatic and non-automatic downstep are distinct processes or can be regarded as the same phenomenon. Downstep is seen as a lowering of the pitch register, relatively to which tones are scaled. The distinction between the two types of downstep arises from differences in their tonal surface structure: In the case of automatic downstep a L(ow) tone intervening H(igh) tones triggers downstep (HLH); in the case of non-automatic downstep the L tone trigger is not realized phonetically (H!H, ! indicates downstep). Our controlled experimental data on downstep in Akan suggest that the degree of lowering is the same in both cases contrary to what has been claimed earlier in the literature.

Keywords: Akan, automatic downstep, non-automatic downstep.

1. INTRODUCTION
This study reports on a comparison of the phonetic realization between automatic and non-automatic downstep in Akan, a two-tone language belonging to the Kwa branch of the Niger-Congo phylum, which is one of the major languages spoken in Ghana [13].

If the two types of downstep are phonetically identical, they share the same phonological trigger (a L tone), and can be analysed as one process. There has been a debate whether to treat the two as alike or as distinct processes (cf. e.g. [17, 18, 19] vs. [7, 15]), and in particular for Akan [7] claims that they are phonetically distinct. On the contrary [19] argues for an alike treatment. The present study will quantitatively address this issue by comparing the two types of downstep in identical sentence frames to avoid any influence from tonal structure or sentence length.

1.1. Previous work on tone in Akan
Akan exhibits a tonal contrast between lexical H and L tones, the tone bearing unit (TBU) is the syllable [6]. According to [6] Akan distinguishes three syllable types (V, CV, C), single sonorant consonants function as syllabic consonants, and any vowel constitutes a syllable, and in case of two adjacent vowels each of them constitutes its own syllable.

Akan has been classified as terraced-level tone language [1, 3, 19], which according to [3] displays a regular process of register shift. The register consists of reference lines relative to which local tonal targets are scaled. The shift of the pitch register affects the scaling of successive tones because it sets a new ceiling for tonal scaling while preserving the underlying phonological distinction.

The terracing property relates to downstep in the way that a H tone following a L tone is realized at a lower pitch level (automatic downstep, which is often called downdrift; see [5] for an overview). In the case of non-automatic downstep, or simply downstep, a similar lowering of H tones occurs, yet the triggering L tone is not phonetically realized. The underlying L tone is dissociated with its segmental material, e.g. by segmental deletion, and is often analysed as a floating tone [10]. An example of non-automatic downstep in Akan is given in (1a), which contrasts phonemically with a non-downstepped instance (1b) [6].

(1) a. ɔbɔ!nɔ - ‘messenger’
  b. ɔbɔ!nɔ - ‘creator’

In (1a) the last H tone is downstepped, and no overt L tone is realized prior to the downstepped H tone. This is analysed as lexical downstep, in which the downstepped H tone is part of the lexical entry of the word see a. o. [6] for Akan. [1] argues that instances of lexical downstep in Akan can be etymologically backtracked to exhibit a floating L. The L tone triggers the following H tone to be lowered, a. o. [4].

If there is a floating L tone involved in the derivation of non-automatic downstep it is not to be treated as lexical downstep as [6] does and we would expect no phonetic difference between the
two types of downstep since they have an identical L tone trigger.

1.2. Downstep in Akan

[7] examined the phonetic realization of automatic downstep (2a) and non-automatic downstep (2b) by recording 5 male speakers of Akan.1 Her material for testing automatic downstep (2a) exhibits an alternating tonal make up of H and L tones. The first H tone (H₁) is preceded by a L tone; however the initial tone should not be affected by automatic downstep since it serves as frame setter [9]. The following H₂ and H₃ tones are subject to automatic downstep. The H tone following H₃ is not separated by a L tone and is therefore assumed to exhibit the same tonal height as H₃.

\[
\begin{align*}
(2) & \text{a. } \text{L H₁ L H₂ L H₃ } \quad \text{HL H₄} \\
& \quad \text{Papa Kofi kòtò ntòma²} \\
& \text{Father Kofi go_buy.PRS cloth} \\
& \text{‘Father Kofi has gone to buy cloth.’} \\
\end{align*}
\]

The sparse literature leaves us with conflicting results for Akan with [19] claiming identical phonetic realization, and [7] promoting a difference with a greater degree of lowering for non-automatic downstep.

1.3. Downstep in other tone languages

Two instrumental studies on Igbo (Kwa) show conflicting results, too. [15] reports that automatic downstep causes a significantly greater degree of lowering than non-automatic downstep. Whereas [12] concludes that both types exhibit the same amount of lowering.

[7] examined the phonetic realization of downstep in Bimoba (Gur). Both L and M tones cause H tones to be downstepped. No phonetic difference between the two types of downstep was observed. Also for Chumburung (Kwa) [18] concludes that the degree of lowering is the same and that his findings support theories which equate automatic and non-automatic downstep [e.g., 9, 16].

2. METHOD

2.1. Speech materials

We embedded two target words (3), causing different downstep types, into the same carrier sentence (4). In the case of automatic downstep (3a) the L tone is present at the surface. In the case of non-automatic downstep (3b) the nominal prefix .Subject is deleted, and the L tone dissociates. The floating L tone triggers downstep on the following H tone.

\[
(3) \text{a. Automatic downstep:} \\
\text{kòfì + pàpà } \rightarrow \text{kòfì pàpà} \\
\text{Kofi father Kofi’s father}
\]

The results are summarized by recording 5 male speakers of Akan.

Table 1. F0 values in Hz and differences in pitch drop for automatic and non-automatic downstep, data from [7].

<table>
<thead>
<tr>
<th>H in Hz</th>
<th>Δ</th>
<th>H in Hz</th>
<th>Δ</th>
</tr>
</thead>
<tbody>
<tr>
<td>H₁</td>
<td>181</td>
<td>31</td>
<td>180</td>
</tr>
<tr>
<td>H₂</td>
<td>150</td>
<td>10</td>
<td>155</td>
</tr>
<tr>
<td>H₃</td>
<td>140</td>
<td>20</td>
<td>135</td>
</tr>
<tr>
<td>H₄</td>
<td>120</td>
<td>115</td>
<td></td>
</tr>
</tbody>
</table>

The underlying L tone of the infinitive béká ‘come to tell’ causes downstep on the following H tone; this…”!H tone spreads to the preceding low tone syllable, that is, tone spreading is left to right.” [7, p.4]. If spreading occurs in this direction the surface form in (2b) cannot be derived. Since [7] talks about spreading to the preceding tone, the spreading direction can only be right to left which is typologically very uncommon [11]. On this H tone, however, the difference between the two types of downstep becomes apparent (cf. table 1). We leave the exact analysis of this kind of downstep open for further research.

1 The speech data of the present study and that of [7] comes from Asante Twi, the major dialect of Akan.

2 Transcription, association lines and glossing added to the original.

3 According to [7] H₂ and H₃ are derived from !H tone spreading, while H₄ is an instance of “lexical downstep”. The underlying L tone of the infinitive béká ‘come to tell’ causes downstep on the following H tone; this…”!H tone spreads to the preceding low tone syllable, that is, tone spreading is left to right.” [7, p.4]. If spreading occurs in this direction the surface form in (2b) cannot be derived. Since [7] talks about spreading to the preceding tone, the spreading direction can only be right to left which is typologically very uncommon [11]. On this H tone, however, the difference between the two types of downstep becomes apparent (cf. table 1). We leave the exact analysis of this kind of downstep open for further research.

The underlying L tone of the infinitive béká ‘come to tell’ causes downstep on the following H tone; this…”!H tone spreads to the preceding low tone syllable, that is, tone spreading is left to right.” [7, p.4]. If spreading occurs in this direction the surface form in (2b) cannot be derived. Since [7] talks about spreading to the preceding tone, the spreading direction can only be right to left which is typologically very uncommon [11]. On this H tone, however, the difference between the two types of downstep becomes apparent (cf. table 1). We leave the exact analysis of this kind of downstep open for further research.

The underlying L tone of the infinitive béká ‘come to tell’ causes downstep on the following H tone; this…”!H tone spreads to the preceding low tone syllable, that is, tone spreading is left to right.” [7, p.4]. If spreading occurs in this direction the surface form in (2b) cannot be derived. Since [7] talks about spreading to the preceding tone, the spreading direction can only be right to left which is typologically very uncommon [11]. On this H tone, however, the difference between the two types of downstep becomes apparent (cf. table 1). We leave the exact analysis of this kind of downstep open for further research.

The underlying L tone of the infinitive béká ‘come to tell’ causes downstep on the following H tone; this…”!H tone spreads to the preceding low tone syllable, that is, tone spreading is left to right.” [7, p.4]. If spreading occurs in this direction the surface form in (2b) cannot be derived. Since [7] talks about spreading to the preceding tone, the spreading direction can only be right to left which is typologically very uncommon [11]. On this H tone, however, the difference between the two types of downstep becomes apparent (cf. table 1). We leave the exact analysis of this kind of downstep open for further research.

The underlying L tone of the infinitive béká ‘come to tell’ causes downstep on the following H tone; this…”!H tone spreads to the preceding low tone syllable, that is, tone spreading is left to right.” [7, p.4]. If spreading occurs in this direction the surface form in (2b) cannot be derived. Since [7] talks about spreading to the preceding tone, the spreading direction can only be right to left which is typologically very uncommon [11]. On this H tone, however, the difference between the two types of downstep becomes apparent (cf. table 1). We leave the exact analysis of this kind of downstep open for further research.
b. Non-automatic downstep:

\begin{align*}
\text{kòfì} + \circ\text{-dán} & \rightarrow \text{kòfì} \: \text{!dán} \\
\text{Kofi} & \quad \text{N-house} \quad \text{Kofi's house}
\end{align*}

\begin{align*}
\text{(4)} & \quad \text{LH}_1 \text{L} \quad \text{LH}_2 \! \text{HH}_3 \quad \text{H} \quad \text{L} \quad \text{H} \\
\text{Afua hunù} \quad \text{Kofi dá ŋ} \quad \text{ánèpá} \quad \text{yì} \\
\text{Afua sec.PRS Kofi house this morning.}'
\end{align*}

2.2. Procedure

Six native speakers of Akan (5 male, 1 female) participated in the experiment. Their average age was 28. All speakers declared English as their second language.

The experiment was carried out, using presentation software, in Akan script; since it does not mark tones an English translation was given below the text. Items from two other unrelated experiments were interspersed as fillers. Participants were digitally recorded at a sampling frequency of 44.1 kHz and 32 bit resolution, on a laptop (Levonon R61) using Audacity (Version 1.2.6) and a headset (Logitech Internet Chat Headset). The microphone was an electret condenser type with a sensitivity of -39 dBV/Pascal.

Each test sentence was repeated three times which results in a dataset of 36 sentences (6 speakers x 2 downstep conditions x 3 repetitions).

2.3. Data pre-processing, statistical analysis

The syllables of all sentences were labelled by hand in Praat [2], and boundaries were set automatically at zero crossings. The F0 analysis was based on a Hanning window of 0.4 seconds length containing either automatic (solid line) or non-automatic downstep (dashed line). The crucial part of the target word, i.e. the second syllable carrying the downstepped H tone, is highlighted in bold and by a box.

The curves are nearly lying upon each other. The automatically downstepped H tone on the second syllable of the target word pàpà is realized at a mean F0 of 140 Hz (SD=36.5). The non-automatically downstepped H tone of the first syllable of the target word dàn exhibits a mean F0 of 140 Hz (SD=27.8). The paired samples T-test did not reveal a significant difference (t(5) = 0.002, p = .998).

The first syllable of the target word dàn defines a new register for H₃ (4). We compare the second syllable of both words to exclude any positional effects. H₃ on the second syllable of the target word dàn, is realized at a mean F0 of 141 Hz (SD=27.8). The paired samples T-test did also not show a significant difference between the two H tones (t(5) = 0.09, p = .932), cf. the box in Fig 1.

To compare our results with those of [7] we calculated pitch drops between the H tones (table 2). Contrary to [7] the amount of drop in pitch is similar between the two types of downstep, and does...
not differ significantly \( t_{H1H2}(5) = 1.3, p = 0.25; t_{H2H3}(5) = 0.5, p = 0.64 \).

4. DISCUSSION

Comparing the phonetic realization of automatic and non-automatic downstep in Akan we show that there is no phonetic difference between the two in line with [19] and in opposition to [7]. In contrast to earlier studies we compared the two types of downstep by embedding target words in an identical sentence frame in order to exclude any influence of sentence length or tonal configuration.

The drop in pitch is generally greater in the beginning of an utterance. This follows from the fact that Akan is a terraced level language in which an utterance is subject to a gradual lowering process [3]. The pitch register decreases over the course of an utterance.

In [7] the pitch drop in the beginning of the automatic downstep case is on average 5 Hz bigger than for non-automatic downstep (cf. table 1). Our calculation of paired samples T-tests for the absolute pitch levels reported in [7] reveals no significant difference between automatic and non-automatic downstep for any of the three H tones \( t_{H1}(4) = 0.6, p = 0.56; t_{H2}(4) = 1.8, p = 0.15; t_{H3}(4) = 2.3, p = 0.083 \).

The second pitch drop between \( H_2 \) and \( H_3 \) is smaller in the case of automatic downstep, which according to our calculation is significant \( t(4) = 7.8, p = 0.001 \). Hence the difference between automatic and non-automatic downste is mainly due to the second drop in pitch; to our view the difference in absolute Hz values is rather due to variation in speech production. The exact reason for the difference in the second drop remains unclear.

However, our study based on controlled data provided evidence that there is only one process of downstep in Akan, which is in line with [12, 17, 18, 19]. The conclusion is based on two measures, (i) the absolute pitch level, and (ii) the pitch drop between H tones.

Though one could be tempted to say that downstep is solely automatic and therefore a phonetic by-product of the tonal configuration there are languages such as English [14] and German [8] where no L tone downstep trigger is present, neither an underlying one nor one on the surface.

Acknowledgements: This paper is part of the project D5 (SFB 632 “Information Structure”) at Potsdam University, financed by the DFG. We express our gratitude to Afua Blay and Lisa Pieplow-Stagg.

5. REFERENCES

Interdisciplinary Studies on Information Structure
Vol. 13

Linguistic Fieldnotes I:
Information Structure in different African Languages

Mira Grubic, Susanne Genzel, Frank Kügler
Mira Grubic | Susanne Genzel | Frank Kügler

Linguistic Fieldnotes I: Information Structure in different African Languages
How to elicit semi-spontaneous focus realizations with specific tonal patterns

Susanne Genzel & Frank Kügler
Universität Potsdam

This article presents a situation description production experiment investigating the interaction between syntax and information structure in Akan, a tone language that belongs to the Kwa branch of the Niger-Congo family spoken in Ghana. Information structure was elicited via context questions that put the object in narrow informational focus or narrow corrective focus while controlling for the tonal structure of the target word. Contrary to the prediction that corrective focus is marked by fronting and morphological marking of the focused constituent the data suggest that the in-situ strategy is the preferred one.

Keywords: Akan, information structure, syntax

1 Introduction

Akan is a tone language exhibiting lexical high (H) and low (L) tones (Dolphyne 1988). It is one of the major languages of Ghana spoken by about 8.3 million people (Lewis (ed.) 2009), belonging to the Kwa languages, within the Niger-Congo phylum (Storch 2001). This paper focuses on Asante-Twi, although throughout the paper we will use the name Akan. This is in order to preclude tonal and segmental differences among the dialects (for a detailed discussion about the differences see Cahill 1985; Dolphyne 1988; Abakah 2002; 2005 and 2010).
Abakah & Koranteng 2007 among others). Akan is a SVO language (Kobele & Torrence 2006). The following example (1) illustrates this.

(1) Kofi bɔ-ɔ\(^1\) ama
    kofi hit-PST\(^2\) ama
    ‘Kofi hit Ama.’
(adapted from Kobele & Torrence 2006, p. 162 gloss adjusted)

A simple clause as in (1) uttered without any further context is considered to be used with information focus. The usage of information structural terminology is inspired by Krifka (2008) distinguishing between information and correction focus. Both can be either wide or narrow, and refer to the corresponding alternatives that the context may evoke. If a constituent appears in its base position we will refer to this construction as in-situ, while the term ex-situ means that the constituent is moved from its base position to a designated focus position. Narrow focused constituents in Akan can be left in-situ, or marked by an ex-situ construction which shows the following characteristics:

(i) left-peripheral dislocation of a constituent
(ii) introduction of a clitic morpheme after the dislocated element
(iii) pronoun resumption in a canonical clause position
(Marfo & Bodomo 2005, p. 180)

If, for example, the direct object in (1) is narrowly focused it moves to the left periphery and is followed by the focus marker \textit{na}. The following example (2) illustrates this.

---

\(^1\) In the examples we will use the Akan orthography as laid out in Dolphyne (1988).
\(^2\) The abbreviations used follow the Leipzig Glossing Rules (Comrie et al. 2008).
All NPs can be pronominalized in object position, the resumptive pronoun is over if the dislocated element is animate and covert if it is inanimate (Saah, 1988). According to Boadi (1974) two particles, *na* and *deε*, may function as the clitic morpheme after the dislocated element. *Na* is characterized by him as exclusive focus marker, which narrows down the referential range of the constituent it attaches to and places it in an exclusive class so that it contrasts with other members of the sets of alternatives (Boadi 1974, p. 7). *Deε* however is not exclusive. Moreover it presupposes a weaker commitment of the speaker towards the new information (see Ameka 2010 for an analysis of *deε* as topic marker). *Na* can be used to mark any constituent whereas *deε* is more distributionally limited (Boadi 1974, p. 9). Concerning the prosodic marking of the ex-situ constructions, Boadi (1974) claims that L toned *na*-focused elements are raised to H and that H toned elements are raised one step higher (Boadi 1974, p. 19).

In Akan, focus-constructions and wh-questions show interesting common characteristics (Saah 1988, p. 26). With respect to wh-questions, Saah (1988) and Kobele & Torrence (2006) report that Akan has two options for wh-placement: the wh-element may appear in-situ (3a) or ex-situ (3b). In the ex-situ variant the fronted question word is also followed by the focus marker *na.*

---

3 Brackets indicate optionality.
(3) a. Kofi bɔ-ɔ hena
kofi hit-PST who
‘Who did Kofi hit?’

b. Hena na o-bɔ-ɔ kofi
who FM PRO-hit-PST kofi
‘Who is it that Kofi hit?’

(adapted from Kobele & Torrence 2006, pp. 165–166 gloss adjusted)

This is interesting for us, because the two options (3a) and (3b) are not equally distributed. The in-situ strategy seems to be more restricted in use. The wh-element has to be moved out of its base position if it is on the right of a negation particle, and/or focus sensitive particles like only, or even, and in imbedded clauses (Kobele & Torrence 2006). Saah (1988) reports a slight semantic difference between the two wh-constructions with ex-situ wh-constructions being more emphatic. A similar claim is made by Ermisch (2006). For the focus construction she reports on a difference in marking between informational focus and identificational focus\(^4\). Informational focus goes along with the in-situ construction and identificational focus is marked by the ex-situ construction and has an exhaustive reading. Marfo & Bodomo (2005) also observe that a constituent cannot be contrastively focused in-situ in Akan (Marfo & Bodomo 2005, p. 187). It is not clear how exhaustive this statement is meant since they only show that in-situ focus with na is ungrammatical.

What becomes evident from this review is that there are two possible constructions for marking narrow focus in Akan, in-situ and ex-situ. Whereas the former is said to occur with informational focus, the latter with corrective focus. The dependence on focus type as put forward by Ermisch (2006) and Saah (1988) will be tested here.

\(^4\) We use the term corrective focus instead of identificational focus (see Krifka 2008).
2 Method

Eliciting information structure can be done in many ways. If the researcher has an advanced knowledge of the language under observation and is familiar with the different focus strategies a simple question-answer design brings good results in the laboratory. The main advance is that the researcher can fully control presuppositions and factors involved. Difficulties arise if the available literature does not exhaustively cover speaker’s preferences of the possible strategies and/or environments of usage of one or the other syntactic strategy as it is the case for Akan. Before one can run a controlled experiment about e.g. the prosodic expression of focus in any language one should know the environment in which it is natural to use the one or the other syntactic strategy.

The Questionnaire on Information Structure (QUIS, Skopeteas et al. 2006) is a powerful tool to test for information structural effects. To kill two birds with one stone we adopted one of the QUIS tasks, the Description of single situation. First we wanted to find out about possible strategies of focus realisation for narrow informational focus and narrow corrective focus applied by the speakers, and which of the two attested is the preferred one for a controlled experiment testing the prosodic expression of focus in Akan (see Kügler & Genzel, re-submitted). And second we wanted to elicit a semi-spontaneous dataset with the same tonal configurations as used in the controlled experiment (see Kügler & Genzel, re-submitted).

2.1 Material

Figure 1 shows the two adapted pictures designed for the Description of a single situation task.
Figure 1 Adopted pictures for the description of a single situation task.

The picture on the left panel in Figure 1 illustrates the situation of the test sentence (4a) and the picture on the right of the sentence (4b).

(4) a. Agyeman boa-a addo anɔpɔ yi
    agyeman help-PST addo morning this
    ‘Agyeman helped Addo this morning.’

  b. Anum tɔ-ɔ amango anɔpa yi
    anum buy-PST mango morning this
    ‘Anum bought mangos\(^5\) this morning.’

To elicit the desired information structure the following pre-recorded questions (5) and (6) were used. The first question (5a) was used for both pictures to make sure that the participants understood the situation displayed (5a = 6a).

(5/6)a. Deebɛn na wo-hunu wo saa mfoni yi mu
    what FM PRO-see.PRS be_at_a_place.PRS this picture this in
    ‘What do you see on this picture?’

The question in (5b) seeks to elicit narrow informational focus on the object. It is a wh-question asking for the object. The third question (5c) seeks to elicit

\(^5\) We translate \textit{amango} as plural ‘mangos’, although there is a plural form \textit{mmango} which is not used in our sample below. In the picture more than one mango is displayed and even if \textit{amango} is used with the numeral \textit{many} the speakers did use it in its singular form.
narrow corrective focus on the object by contrasting it with another name in the question.

(5) b. Hwan na agyeman boa-a anɔpa  yi
   who  FM agyeman help-PST morning this
   ‘Whom did Agyeman help this morning?’

c. Agyeman boa-a anum anɔpa  yi
   agyeman help-PST anum morning this
   ‘Did Agyeman help Anum this morning?’

The question in (6b) elicits narrow informational focus on the object in the picture on the right panel in Figure 1. Question (6c) elicits narrow corrective focus on the object by contrasting it with another comestible good in the question.

(6) b. ɛdeɛn na anum tɔ-ɔ anɔpa  yi
   what  FM anum buy-PST morning this
   ‘What did Anum buy this morning?’

c. Anum boa-a kɔbi anɔpa  yi
   anum buy-PST salty fish morning this
   ‘Did Anum buy salty fish this morning?’

2.2 Procedure

The pictures illustrating the situations (Figure 1) were presented to the participants on paper. They were asked to answer the pre-recorded wh-questions (see (5)–(6) above) in a natural way, using full sentences. The participants were informed that the whole situation happened this morning and were instructed to use the temporal information when answering the questions.
The questions were spoken by a young female native speaker and were recorded in a quiet room in Berlin directly on a laptop (Levono R61) using Audacity (Version 1.2.6) and a headset (Logitech Internet Chat Headset). The microphone was an electret condenser type with sensitivity of -39 dbV/Pascal.

The participants heard the pre-recorded questions over headphones. The headphones were binaural with a frequency spectrum from 20-20000 Hz and an acoustic impedance of 32 Ohm with an integrated volume control, so that every participant could adjust the volume. The answers were digitally recorded on a laptop (Levono R61) using Audacity (Version 1.2.6) in a quiet room in the Linguistics Department at the University of Ghana using the same headset.

2.3 Participants

Eleven native speakers of Akan (6 female and 5 male) as spoken in and around Kumasi participated in the experiment. Eight participants were students of the University of Ghana in Accra. For the other three, one was doing his national service in Kumasi, another working in the pharmacy, and the third one working as university lecturer. The average age was 26 years. All speakers declared English as their second language. Each speaker was paid a small fee for participation.

3 Results

In this section we list all answers given by the participants with the number of the question first and the number of the participant as second digit separated by a dot (e.g. (5a.9) is the answer to question 5a from speaker 9).
(5a.1) Me-hunu se agyeman re-boa a-yi addo afiri PRO-see.PRS that agyeman PROG-help PRO-take_out.PRS addo from nsuo no mu anɔpa yi water the out morning this ‘I see that Agyeman is helping to take Addo out of the water this morning.’

(5b.1) Agyeman boa-a addo anɔpa yi agymean help-PST addo morning this ‘Agyeman helped Addo this morning.’

(5c.1) Daabi agyeman boa-a addo anɔpa yi no agymean help-PST addo morning this ‘No! Agyeman helped Addo this morning.’

(5a.2) Me-hunu agyeman ene addo PRO-see.PRS agyeman and addo ‘I see Agyeman and Addo.’
Me-hunu se addo ε-wɔ nsuo emu PRO-see.PRS that addo PRO-be_at_a_place.PRS water in ‘I see that Addo is in the water.’
Me-hunu se agyeman nso e-te kodo bi mu PRO-see.PRS that agyeman also PROG-sit boat one in ‘I see that Agyeman is sitting in a boat.’
Addo atene ne nsa ma agyeman e-fa no ε-si addo hold.PRS PRO hand out agyeman PROG-take PRO PRO-stand.PRS kodo bi no mu boat the in ‘Addo holds out his hand for Agyeman for taking him, he stands in the boat.

(5b.2) Agyeman boa-a addo agymean help-PST addo ‘Agyeman helped Addo.’

(5c.2) Daabi addo na agyeman boa-a no anɔpa yi no addo FM agyeman help-PST PRO morning this ‘No! It was Addo who Agyeman helped this morning.’
(5a.3) Me-hunu së agyeman ε-te kodoc mu
PRO-see.PRS that agyeman PRO-sit.PRS boat in
‘I see that Agyeman sits in a boat.’
Na addo nso da nsuo mu a agyeman pëse
and addo also lie.PRS water in and agyeman want.PRS
ɔ-boa no na ɔ-yi no firí nsuo no mu
PRO-help.PRS PRO and PRO-take_out.PRS PRO from water the out
‘And Addo also lies in the water and Agyeman wants to take him
out of the water.’

(5b.3) Agyeman boa-a addo anɔpa yi
agyemean help-PST addo morning this
‘Agyeman helped Addo this morning.’

(5c.3) Daabi agyeman boa-a addo anɔpa yi
no agyemean help-PST addo morning this
‘No! Agyeman helped Addo this morning.’

(5a.4) Me-hunu së akɔdab bi a ye-fre no addo ɔ-da
PRO-see.PRS that child one who PASS-call PRO addo PRO-lie.PRS
nsuo mu
water in
‘I see that a child who is called Addo lies in the water.’
Nsuo de no ε-ko na agyeman
water do_sth_with.PRS PRO PRO-drawn.PRS and agyeman
ε-pë
PRO-want.PRS
ɔ-boa na yi no firí nsuo no mu
PRO-help.PRS and take_out.PRS PRO from water the out
‘The water wants to drawn him and Agyeman wants to help him and
take him out of the water.’

(5b.4) Agyeman boa-a addo
agyemean help-PST addo
‘Agyeman helped Addo.’

(5c.4) Daabi ɔ-boa-a addo
no PRO-help-PST addo
‘No! He helped Addo.’
(5a.5) Me-hunu sɛ addo a-tɔ nsuo mu na agyeman pɛɛɛ
PRO-see.PRS that addo PERF-fall water in and agyeman want.PRS
ɔ-yi no firi mu
PRO-take_out.PRS PRO from out
‘I see that Addo fell in the water and Agyeman wants to take him out.’

(5b.5) Agyeman boa-a addo anɔpa yi
agymean help-PST addo morning this
‘Agyeman helped Addo this morning.’

(5c.5) Daabi agyeman boa-a addo anɔpa yi
no agymean help-PST addo morning this
‘No! Agyeman helped Addo this morning.’

(5a.6) Me-hu m-barima mmienu
PRO-see.PRS PL-man two
‘I see two men.’
Agyeman te suhyɛma mu w-a-tene ne nsa pɛ
agymean sit.PRS boat in PRO-PERF-stretch PRO hand want.PRS
sɛ ɔ-sɔ addo mu addo da nsuo no mu
that PRO-hold.PRS addo in addo lie.PRS water the in
‘Agyeman sits in a boat, his hands are stretched, he wants to
hold Addo, Addo lies in the water.’

(5b.6) Agyeman boa-a addo
agymean help-PST addo
‘Agyeman helped Addo.’

(5c.6) Mepawokyɛw daabi ɔ-m-boa-a anum
please no PRO-NEG-help-PST anum
‘Please no! He did not help Anum.’
Me-hu sɛ addo ε-da nsuo mu na agyeman gyina PRO-see.PRS that addo PRO-lie.PRS water in and agyeman stand.PRS kodoɔ mu nansɔ nsuo re-fa addo nti agyeman pɛ sɛ boat in but water PROG-take addo so agyeman want.PRS that ø-twe no a-ba kodoɔ no mu PRO-pull.PRS PRO PERF-come boat the in ‘I see that Addo lies in the water and Agyeman stands in the boat but the water is taking Addo so Agyeman wants to pull him to come in the boat.’

Addo addo ‘Addo!’

Daabi daabi ε-n-yɛ addo na ø-boa no no no PRO-NEG-be.PRS addo FM PRO-help.PRS PRO ‘No! No! It is not Addo whom he helps!’

Me-hu sɛ addo da nsuo mu agyeman ø-te hyɛn PRO-see.PRS that addo lie.PRS water in agyeman PRO-sit.PRS ship mu a ø-pɛ sɛ ø-yi no firi nsuo no mu in who PRO-want.PRS that PRO-take_out.PRS PRO from water the in ‘I see that Addo lies in the water Agyeman sits in a ship he wants to take him out of the water!’

Agyeman boa-a addo anɔpa yi agymean help-PST addo morning this ‘Agyeman helped Addo this morning.’

Daabi agyeman boa-a addo anɔpa yi no agymean help-PST addo morning this ‘No! Agyeman helped Addo this morning.’
(5a.9) Saa mfoni yi mu ɔbarima ye-frɛ no agyeman a ɔ-te
this picture this in man PASS-call PRO agyeman who PRO-sit.PRS
ɛhyɛn bi mu
ship one in
‘In this picture: a man called Agyeman who is in a ship.’
W-a-tene ne nsa de re-ma abofra bi a ye-frɛ
PRO-PERF-stretch PRO hand for PROG-give child one who PASS-call
no addo
PRO addo
‘He has stretched his hand for giving it to a child who is called
Addo.’
Addo da nso mu na ɔnɔ nso a-tene ne nsa de
addo lie.PRS water in and PRO also PERF-stretch PRO hand for
a-ma agyeman
PERF-give agyeman
‘Addo lies in the water and he has also stretched his hands for
giving it to Agyeman.’

(5b.9) Anɔpa yi agyeman boa-a abɔfra bi a ye-frɛ no addo
Morning this agyeman help-PST child one who PASS-call PRO addo
‘This morning Agyeman helped a child who is called Addo.’

(5c.9) Daabi agyeman boa-a addo
no agyeman help-PST addo
‘No! Agyeman helped Addo.’

(5a.10) Me-hu sɛ addo da nso no mu na agyeman nso
PRO-see.PRS that addo lie.PRS water the in and agyeman also
ɛ-te kodoɔ no mu na w-a-tene ne nsa pe
PRO-sit.PRS boat the in and PRO-PERF-stretch PRO hand want.PRS
sɛ ɔ-sɔ agyeman nso nsa
that PRO-hold.PRS agyeman also hand
‘I see that Addo lies in the water and Agyeman sits in the boat also
and he has stretched his hand wanting to hold Agyeman’s hand too.’

(5b.10) Agyeman boa-a addo anɔpa yi
agyeman help-PST addo morning this
‘Agyeman helped Addo this morning.’
(5c.10) Daabi agyeman boa-a addo anɔpa yi
no agyeman help-PST addo morning this
‘No! Agyeman helped Addo this morning.’

(5a.11) Mfonì yi μu me-hu se addo a-tɔ nsuo no μu a
Picture this in PRO-see.PRES that addo PERF-fall water the in who
neho re-kyere no himself PROG-point_out PRO
‘In the picture I see that Addo has fallen in the water he is calling for his
attention.’
Agyeman a-tene ne nsa se o-re-sɔ ne μu na
agyeman PERF-stretch PRO hand that PRO-PROG-hold PRO in and
w-boa no a-ma n-a-foro kodoɔ no sɛɛɛɛ PRO-help.PRS PRO PERF-give PRO-PERF-climb boat the so
ε-бе-yɛ o-be-nya nkwa sɛɛɛɛ ε-бе yɛ nsuo no
PRO-FUT-be PRO-FUT-get life so PRO-FUT-be water the
m-fa no NEG-take.PRS PRO
‘Agyeman has stretched his hand to drip him and help him climb into the
boat so that the water will not take him.’

(5b.11) Agyeman boa-a addo anɔpa yi
agyeman help-PST addo morning this
‘Agyeman helped Addo this morning.’

(5c.11) Daabi agyeman boaa addo anɔpa yi
no agyeman help-PST addo morning this
‘No! Agyeman helped Addo this morning.’
(6a.1) Me-hunu sɛ anum re-tɔ amango wo
PRO-see.PRS that anum PROG-buy mango be_at_a_place.PRS
mfoni no mu
picture the in
‘I see that Anum is buying mangos in the picture.’

Me-hunu ɔba a ɔ-tɔn amango wo
PRO-see.PRS woman who PRO-sell.PRS mango be_at_a_place.PRS
mfoni no mu
picture the in
‘I see a woman who is selling mangos in the picture.’

Me nso hunu dua wo mfoní no mu
PRO also see.PRS tree be_at_a_place.PRS picture the in
‘I also see a tree in the picture.’

Me-hunu sɛ amango gu ɛponɔ so wo  mfoní
PRO-see.PRS that mango on table much be_at_a_place.PRS picture
no mu
the in
‘I see a lot of mangos on the table in the picture.’

Me-hunu sɛ anum a-ma ne nsa so
PRO-see.PRS that anum PERF-lift PRO hand up
‘I see that Anum has lifted his hands up.’

(6b.1) Anum tɔ-ɔ amango anɔpa yi
anum buy-PST mango morning this
‘Anum bought mangos this morning.’

(6c.1) Daabí Anum tɔ-ɔ amango anɔpa yi
no anum buy-PST mango morning this
‘No! Anum bought mangos this morning.’
(6a.2) Me-hunu  abrantε bi
PRO-see.PRS boy one
‘I see a boy.’
Me-hunu  amango
PRO-see.PRS mango
‘I see mangos.
Me-hunu  maame bi nso
PRO-see.PRS woman one also
‘I also see a woman.’
Afei me-hunu  εpονε sε  amango gu so
now PRO-see.PRS table that mango on much
‘Now I see a table with a lot of mangos on it.’
Me-hunu  dua bi nso sε  ε-wo
PRO-see.PRS tree one also that PRO-be_at_a_place.PRS
ahaban  bebere
leaves many
‘I see also a tree with many leaves.’
Me-hunu  kenten bi  sε  ε-si
PRO-see.PRS basket one also that PRO-be_situated.PRS
εpονε  ase
table under
‘I see also a basket under the table.’
Me-hunu  sε  abrantε no nso kura  sika
PRO-see.PRS that boy  PRO also carry.PRS money
‘I see that a boy carries money.’

(6b.2) Anum  tɔ-ɔ  amango
anum  buy-PST mango
‘Anum bought mangos.’

(6c.2) Daabi Anum  tɔ-ɔ  amango anɔpa  yi
no anum  buy-PST mango morning this
‘No! Anum bought mangos this morning.’
(6a.3) Me-hunu barima bi a y-akyere natadεε addo mu
PRO-see.PRS man one who PASS-write shirt addo in
‘I see a man, his shirt has the inscription Addo.’
Me-hunu se maame bi nso re-tɔn amango
PRO-see.PRS that woman one also PROG-sell mangos
‘I also see that a woman is selling mangos.’
Na barima no ε-ye se ɔ-pεɛɛ ɔ-tɔ mango
And man the PRO-be.PRS that PRO-want.PRS PRO-buy.PRS mango
no bi the some
‘And it is the man that wants to buy some mangos.’
Me-hunu adua nso se e-si maame no akyi
PRO-see.PRS tree also that PRO-stand.PRS woman the behind
‘I see also a tree that stands behind the woman.’
Adua no ε-ye ahahanmono
tree the PRO-be.PRS green
‘The tree is green.’

(6b.3) Anum tɔ-ɔ amango anɔpa yi
anum buy-PST mango morning this
‘Anum bought mangos this morning.’

(6c.3) Daabi anum tɔ-ɔ amango anɔpa yi
no anum buy-PST mango morning this
‘No! Anum bought mangos this morning.’

(6a.4) Me-hunu se maame bi ε-tɔn amango ena akɔdaa bi
PRO-see.PRS that woman one PROG-sell mango and child one
gyina ne nyκεν ε-tɔ bi
PROG-stand⁶ PRO nearby PRO-buy.PRS some
‘I see that a woman is selling mangos and a child is standing
nearby her, he wants to buy some mangos.’

(6b.4) ɔ-tɔ-ɔ amango
PRO-buy-PST mango
‘He bought mangos.’

⁶ Aspect is marked by a tonal shift (Dolphyne 1988, pp. 67–68).
(6c.4) Daabi ɔ-tɔ-ɔ amango wa-n-tɔ kobi
no PRO-buy-PST mango PRO-NEG-buy.PRS salty fish
‘No! He bought mangos not salty fish.’

(6a.5) Me-hunu se anum gyina maame bi a ɔ-tɔn
PRO-see.PRS that anum PROG-stand woman one who PRO-sell.PRS
amango ho mango aside
‘I see that Anum is standing next to a woman who sells mangos.’
Maame no ɔ-tɔn amango no wɔ dua ase
woman the PRO-sell.PRS mangos the be_at_a_place.PRS tree under
‘The woman sells mangos under the tree.’

(6b.5) Anum tɔ-ɔ amango anɔpa yi
anum buy-PST mango morning this
‘Anum bought mangos this morning.’

(6c.5) Daabi anum tɔ-ɔ amango anɔpa yi
no anum buy-PST mango morning this
‘No! Anum bought mangos this morning.’
(6a.6) Me-hu abaamoa bi ye-frε no anum c-gyina maame PRO-see.PRS boy one PASS-call PRO anum PRO-stand.PRS woman bi nkyεn one side ‘I see a boy called Anum he stands next to a woman.’
Maame no tɔn amango na c-pe se c-tɔ woman the sell.PRS mango and PRO-want.PRS that PRO-buy.PRS amango no bi mango the some ‘The woman sells mangos and he wants to buy some mangos.’
Maame no te dua bi ase na w-a-yehye amango woman the sit.PRS tree one under and PRO-PERF-arrange mango no wo c epɔnɔ no so the be_at_a_place.PRS table the top ‘The woman sits under a tree and has arranged the mangos on top of the table.’
Abaamoa no ye-frε no anum no a-pegya ne nsa se boy the PASS-call PRO anum PRO PERF-lift PRO hand that c-re-kyea maame no a c-tɔn amango no PRO-PROG-greet woman the who PRO-sell.PRS mango the ‘The boy called Anum has lifted his hand for greeting the woman who sells the mangos.’
Me-hu se anɔpa yi anum re-kɔtɔ amango PRO-see.PRS that morning this amango PROG-go_buy mango wo dua bi ase be_at_a_place.PRS tree one under ‘I see that this morning Anum goes to buy mangos under the tree.’

(6b.6) Anum tɔ-c amango anum buy-PST mango ‘Anum bought mangos.’

(6c.6) Daabi a-n-tɔ-c kobi no PRO-NEG-buy-PST salty fish ‘No! Anum did not buy salty fish.’
(6a.7) Me-hu së ançpa yi na anum ɛ-re-kɔɛ amango
PRO-see.PRS that morning this PST anum PRO-PROG-go_buy mango
wo dua bi ase
be_at_a_place.PRS tree one under
‘I see that this morning, Anum was going to buy mangos under a tree.’

(6b.7) Deɛ me-hu ɛ-ye amango
what PRO-see.PRS PRO-be.PRS mango
‘What I see are mangos.’

(6c.7) Daabi ɔ-tɔ-ɔ amango
no PRO-buy-PST mango
‘No! He bought mango.’

(6a.8) Me-hu së ançpa yi anum re-ɔ-kɔɛ amango
PRO-see.PRS that morning this anum PROG-PST-go_buy mango
wo dua bi ase
be_at_a_place.PRS tree one under
‘I see that this morning Anum is going to buy mangos under the tree.’

(6b.8) Anum tɔ-ɔ amango ançpa yi
anum buy-PST mango morning this
‘Anum bought mangos this morning.’

(6c.8) Daabi Anum tɔ-ɔ amango ançpa yi
no anum buy-PST mango morning this
‘No! Anum bought mangos this morning.’
(6a.9) Anum hye hyeti akokɔsradeɛ
anum wear.PRS shirt yellow
‘Anum wears a yellow shirt.’
Maame no nso ɛ-fra ntoma
woman the also PROG-dress cloth
‘The woman is also dressed in cloth.’
Kɔkɔɔ kakra wɔ mu na w-a-san a-bo
red little be_at_a_place.PRS in and PRO-PERF-also PERF-create
duku
‘There is a little bit of red inside and she also has wrapped a bandana.’
Wɔ-gyina dua ahabanmono bi ase
PRO-stand.PRS tree green one under
‘They stand under a green tree.’

(6b.9) Anɔpa yi anum tɔ-ɔ amango
morning this anum buy-PST mango
‘This morning Anum bought mangos.’

(6c.9) Daabi anɔpa yi anum tɔ-ɔ amango
no morning this anum buy-PST mango
‘No! This morning Anum bought mangos.’

(6a.10) Me-hu abrantee bi ye-fre ɛ no anum ɔ-pe se
PRO-see.PRS boy one PASS-call PRO anum PRO-want.PRS that
ɔ-tɔ amango
PRO-buy.PRS mango
‘I see a boy called Anum he wants to buy mangos.’
Afei nso me-hu maame bi a ɔ-tɔn amango
again also PRO-see.PRS woman one who PRO-sell.PRS mango
‘Again I see also a woman who sells mangos.’
Amango pii nso gu ɛpon no so
mango many also on table the top
‘Also many mangos lie on top of the table.’
Saa nso na me-hu dua bi wɔ maame no
this also and PRO-see.PRS tree one be_at_a_place.PRS woman the
akyi
back
‘This too and I see a tree behind the woman.’
First of all the answers to questions (5a) and (6a) show that most of the participants conceived the situation displayed correctly. In the sample (6a.9) the participant does not describe the action displayed, emphasis is put on the description of cloth and landscape. Nonetheless the participant answered the follow up questions as expected.

Concerning the information structural marking the results show that in case of narrow informational focus out of 22 answers no usage of the ex-situ construction could be observed. There is remarkably little variation in the data. We find one single word utterance, (5b.7), repeating only the questioned object,
two other constructions (6b.7) and (6b.10) but with no sign of fronting and/or morphological focus marking. Participant 7 uses a question word followed by the verb *see* and the questioned object, in the sample of participant 10 we see that the questioned object appears in the left periphery of the sentence but without the designated focus maker *na*, followed by a locative construction. Furthermore we find one instance of pronominalization (6b.4).

In the context of corrective focus three out of 22 answers are realized ex-situ (5c.2, 5c.7, 6c.11). Concerning variation we find two instances of pronominalization (5c.4, 6c.7) two participants used a negation on the verb (5c.7, 6c.6) and two combined pronominalization and negation on the verb (5c.6, 6c.4).

4 Discussion

We wanted to test the interaction of information structure and syntactic structure in Akan. Inspired by the work of Ermisch (2006) we hypothesized that narrow informational focus is not syntactically marked i.e. the focused constituent remains in-situ (see also Hartmann & Zimmermann 2007 for comparable results in Hausa). The results of the situation description task support Ermisch’s claim. What is more surprising is that the in-situ strategy is the preferred one, even with narrow corrective focus contra Marfo & Bodomo (2005) and Ermisch (2006). This does not mean that there are no linguistic means of focus marking at all in case of in-situ focus. In our study on the prosodic marking of focus in

---

7 One reviewer asked for question-answer congruency, which is also known as syntactic priming (see Bock 1986). We do not see any influence of syntactic priming, the wh-question asking for an informational focus shows a fronted wh-word which is followed by the focus marker *na*. The question eliciting corrective focus is an in-situ construction. If syntactic priming would matter here, we would have expected more ex-situ answers for informational focus, but the reverse was the case.
Akan (Kügler & Genzel re-submitted) we show contra Boadi (1974) that corrective focus is prosodically marked by means of register lowering on the corrective focused element and on subsequent post-focus constituents regardless of tonal specification. Narrow informational focus remains also prosodically unmarked.

**Reference**


Ermisch, Sonja. 2006. Focus and Topic constructions in Akan. In *Focus and
Semi-spontaneous focus realizations with specific tonal patterns


Section V
A typological overview
Introduction: Tone and intonation from a typological perspective

Sabine Zerbian a,*, Laura J. Downing b, Frank Kügler c

a Department of Linguistics, University of the Witwatersrand, South Africa
b Centre for General Linguistics (ZAS), Berlin, Germany
c Institute of Linguistics, University of Potsdam, Germany

Accepted 23 October 2007
Available online 15 May 2008

Traditionally, tone languages and intonation languages are considered to be typologically distinct. Tone languages use pitch changes to distinguish morphemes, while intonation languages use meaningful pitch changes at the sentence level, often anchoring them to positions of phrasal stress (Ladd, 1996). However, tone and intonation have come to be considered more and more as overlapping areas of research since Pierrehumbert’s (1980) influential Ph.D. dissertation persuasively argued that intonation melodies are best represented as sequences of tones, using the autosegmental formalism developed for the analysis of tone languages. Modern textbooks (Ladd, 1996; Yip, 2002; Gussenhoven, 2004) recognize this fundamental compositional similarity between tone and intonation. Recent research shows, moreover, that there is no clear-cut distinction between tone and intonation languages. Tone languages also make use of intonation, defined as meaningful alternations in pitch across the sentence (e.g., Xu, 1999 for focus in Chinese, Downing et al., 2004 for Chichewa and Rialland (this issue)). Similarly, recent research argues that intonational melodies show tonal behavior in some non-tonal languages (Féry and Kügler, 2008). For reasons like these, Jun (2005) argues for a more fine-grained feature set to classify languages according to how they use pitch changes, placing them along a typological continuum rather than in absolute classes like tone vs. intonation language. For instance, the tone language Mandarin Chinese has both lexical pitch (i.e., tone) and stress, whereas Cantonese only has lexical pitch (Jun, 2005:444). Mandarin thus shares the property of having stress with well-known intonation languages like German or English, even though it is also clearly a tone language (i.e., a language with phonemic tone) while English and German are not. Some of the articles in this
issue present further examples of the kinds of overlap that show why tone and intonation languages cannot easily be separated.

The contributions to this Special Issue were originally presented as talks at the 2nd TIE conference on tone and intonation, organized by the ZAS and University of Potsdam and held at the ZAS in Berlin in September 2006. The TIE conference series grew out of a European network which took as its basic premise that tone and intonation are related phenomena. The network intended to promote research and the exchange of information into topics such as the interaction between intonation and tone, the description of tone and intonation in lesser known languages and dialects, phonetic and perceptual aspects of tone and intonation, the evolution of pitch accent and tonal systems, and tone typology. The articles in this special issue address several of these issues, drawing on data from typologically different languages, spoken in Europe, Asia and Africa. The papers in this issue are grouped into three specific research areas: tone–voicing interactions, phonetic aspects of tone and intonation, and pragmatic aspects of tone and intonation. The contribution of each paper to current debates in these research areas is discussed in turn in the next sections.

1. Tone–voicing interactions

The articles by Hyslop (“Kurtöp tone: a tonogenetic case study”) and Pearce (“Kera tone and voicing interaction”) address the interaction of tone and consonant voicing in two underdescribed languages: Kera, a Chadic language, spoken in Chad and Cameroon, and Kurtöp, a Tibeto-Burman language of Bhutan. It is well established that consonant voicing can interact with tone both diachronically, in motivating tonogenesis (Haudricourt, 1954), and synchronically, in neutralizing tonal contrasts or blocking the spread of tones (Peng, 1992; Bradshaw, 1999). The generalization that emerges from these surveys is that there is a cross-linguistic correlation between voiced consonants and low tone, and between voiceless consonants and high tone. Work like Hombert (1978) provides a phonetic explanation for this correlation. Even in non-tonal languages like English, pitch is raised in a vowel following voiceless consonant and lowered following a voiced consonant. Work like Massaro and Cohen (1976, 1977), cited in Peng (1992), shows that pitch and VOT also reinforce each other as independent perceptual cues to consonant voicing, even for non-tonal languages like English. Their experiments demonstrate that when VOT does not unambiguously identify voicing quality of an obstruent, a low F0 on a following vowel leads listeners to identify the consonant as voiced. Hombert (1978) proposes the following scenario for tonogenesis based on these phonetic studies. As a first stage, voicing would be contrastive for obstruents, and the pitch differences on the vowels following the consonants would provide only a low level secondary cue to voicing. The next step would be for tone to become a more salient cue to voicing, but at this stage, tone would not be contrastive. Instead, syllables beginning with voiced obstruents would be predictably low toned, while syllables beginning with voiceless obstruents (or sonorants, which do not contrast for voice) would be high toned. The final stage of development is for the voicing contrast to be lost, replaced by a tonal contrast. The two endpoints of tonogenesis are clearly illustrated in dialects of Kammu (Svantesson and House, 2006:310), a Mon-Khmer language spoken in northern Laos. Comparing Eastern Kammu with the cognate words in Northern and Western Kammu shows that a voicing contrast in Eastern Kammu has been replaced with a tonal one in the other dialects: Eastern Kammu taaŋ vs. Northern Kammu tāañ ‘pack’; Eastern Kammu daaŋ vs. Northern Kammu tāañ ‘lizard’. As Gussenhoven (2004) points out, a final factor in tonogenesis is language contact. Tone is
mostly an areal feature, shared by unrelated languages that are spoken in close geographic proximity.

The papers by Hyslop and Pearce contribute to our understanding of tone–voice interactions by describing and analyzing languages at a midpoint in the tonogenesis continuum described above. These languages are currently undergoing change in the relative contrastive status of tone and voice, at least partially under the influence of language contact. Hyslop’s study of incipient tonogenesis in Kurtöp shows that tone is only partly contrastive in this language. High and low tone contrast following sonorants and the (voiceless) palatal fricative. However, tone is predictably high following other voiceless obstruents and low following voiced obstruents: the expected pattern at the midpoint of tonogenesis. One question that Hyslop’s paper addresses is why contrastive tone should be entering the language following sonorants and the palatal fricative. Hyslop persuasively argues that the sonorants which are followed by a high tone synchronically were voiceless at an earlier stage of the language, as comparison with cognates in Classical Tibetan shows they were originally preceded by (and plausibly devoiced by) tautosyllabic s. After s was deleted, voicing might have been contrastive on sonorants for some time, but now a tone contrast has replaced the voicing contrast. A former voicing contrast for the palatal fricative has similarly been replaced by a tone contrast. Another question raised by Hyslop’s study is whether Kurtöp shows any signs of losing voicing contrasts in other obstruents besides the palatal fricative, replacing them with tonal contrasts. A pilot phonetic study presented in the paper suggests this could be the case. The production of two speakers, representing two different generations, was analyzed. The results show that the voicing contrast is unreliably realized by both speakers and that the younger speaker shows a stronger tendency not to realize a voicing contrast. Hyslop also proposes that this trend towards tonogenesis is reinforced under the influence of neighboring tonal languages, as we would expect given the proposal that contact is often an important factor in the development of tone systems.

Pearce’s paper shows that Kera is also at a midpoint in the tonogenesis continuum, as tone also appears to be only partly contrastive in this language. As in Kurtöp, tones contrast following sonorants. (Kera has a three-tone system.) However, only low tones are found on vowels following what are transcribed as voiced obstruents, while voiceless obstruents can be followed by both high and mid tones. As Pearce notes, it is accepted by Chadic scholars that the current partially contrastive tonal system is the product of tonogenesis. However, Pearce’s very thorough study provides an important new perspective on the correlation of tone and voice in Kera by showing that tone is, in fact, contrastive following obstruents. It is (the perception of) voicing which is conditioned by tone and not vice versa. That is, in contrast to a language like English, where VOT is a primary cue to voicing and pitch a minor cue, in Kera, pitch is the primary cue to voicing and VOT is a minor cue. Pearce demonstrates that low tone spread in fact conditions what had been previously described as a process of voicing harmony in the language: when a low tone spreads to a suffix, the preceding consonant is perceived as voiced. The suffixal consonant is realized as voiceless when it is followed by a non-low tone. It is exceedingly rare for tone to condition voicing (Yip, 2002; Hombert, 1978; Hansson, 2004), and this paper makes the important contribution of carefully documenting an especially convincing case. Another important contribution of the paper is to document dialect differences in the relative importance of VOT and pitch in signaling voicing. The striking result of this study is to show that VOT plays more of a role in the production and perception of voicing for speakers living in town, where they have more contact with French, than for speakers living in the village. This provides another example where contact is an important factor in motivating change in the laryngeal (tone–voice)
system of a language. As Pearce points out, the intricate interaction of VOT and pitch with the consonant system are difficult to model in theories which adopt a ‘single source’ representation of tone/voice like that proposed by Bradshaw (1999). In these theories, [voice] and low tone are phonetic manifestations of the same feature. This formalism neatly accounts for the fact that low tone spread can result in voicing of a preceding consonant. However, it does not easily explain why VOT and F0 have asymmetric salience in signaling voicing, and why the relative importance of F0 in signaling voicing might change over time. If low tone and voice are manifestations of the same feature, Pearce argues, we might expect F0 to remain stable as a cue to voicing, and this is not what is found in Kera. (See Peng, 1992 for other critiques along these lines of the single source theory.)

2. Phonetic aspects of tone and intonation

The timing of tonal events has been a central issue in tone and intonation research and theory since Bruce’s (1977) analysis of the phonetic correlates of word accents in Central Swedish demonstrated that Accent I and Accent II in Swedish are distinguished primarily by the timing of the word accent fall. These timing issues for Swedish word accents have since been further pursued for different varieties of Swedish in work like Bruce and Gärding (1978) and Bruce and Thelander (2001), who show that dialects that employ the Accent I and Accent II distinction differ in their respective timing of the word accentual tonal gesture.

More recent work on the phonetic aspects of tonal alignment in intonational phonology has pursued a variety of different goals. One central issue in the 1990s has been to demonstrate that tonal targets, rather than tonal movements, are the primitives of intonational description. Work like Ladd et al. (1999, 2000) have shown that measurements of tonal alignment in different speaking styles, for instance, provide evidence for the tonal targets theory, thus solving the ‘levels’ vs. ‘configurations’ debate (cf. Bolinger, 1951; Ladd, 1996).

From a phonological point of view, tonal alignment has been used to test the validity of proposed tonal categories (Pierrehumbert and Steele, 1989; Kohler, 1991; Arvaniti et al., 1998; Ladd and Schepman, 2003). This is possible due to the invariance hypothesis: a tone in a given phonological category shows a relatively stable alignment at certain segmental landmarks of a text. A growing body of descriptive work focuses both on detailed interpretation of phonological tonal categories within single languages (e.g., Gilles, 2005; Kügler, 2004, 2007) and on variation among language varieties (e.g., Atterer and Ladd, 2004). A further line of research on the timing of tonal interactions in tone languages has led to development of the target approximation model (Xu, 2005), which addresses the issue of detecting the basic underlying articulatory processes involved in the production of tone, with the aim of understanding tonal characteristics in more detail and of interpreting surface variation in a more strictly linguistic way.

Accent implementation not only involves the timing of the tonal target but also tonal scaling issues (Kügler, to appear). It remains controversial whether pitch range effects should be interpreted as phonologically categorical or gradient (Ladd, 1994, 1996; Hayes, 1994; Ishihara and Féry, 2006; Féry and Kügler, 2008). The scaling of tones, however, helps to determine whether pitch movements are to be considered tones or mere F0 transitions (Ladd and Schepman, 2003). It is thus a diagnostic of tonal events.

The three papers in this section contribute to different aspects of these issues. While the paper by Prieto uses alignment measurements to argue that bitonal pitch accents of different shapes (i.e., rising or falling) have similar tonal behavior, the paper by Grice et al. manipulates pitch range to argue for floating tones in intonational modeling. Finally, the paper by Xu provides new
insights into the very nature of the timing of tones, distinguishing between obligatory and informational timing.

Prieto’s paper, “Tonal alignment patterns in Catalan nuclear falls,” contributes to the understanding of the nature and behavior of tones in an intonation language. By considering the articulatory coordination of laryngeal and supra-laryngeal gestures, the article provides new evidence on the nature of falling accents. Prieto argues that falling tones behave similarly to (well-studied) rising tones in terms of alignment. Based on the results of an analysis of falling accents in Catalan, Prieto concludes that the first tone of a bitonal accent shows stable alignment with the onset of the stressed syllable, and the second tone of a bitonal pitch accent shows more variable alignment. The speech material consists of recordings of 10 Central Catalan speakers reading ten interrogative sentences five times. The falling accents are elicited with yes–no questions, which in Catalan show a distinct tonal pattern of H+L* pitch accents (Prieto et al., 2007). Much work on the phonetic alignment of tonal targets has been concerned with rising pitch accents (Arvaniti et al., 1998, 2000; Ladd et al., 1999, 2000), while only a limited number of studies and data are available for falling accents. On the basis of these results, Prieto claims that the alignment properties of falling tones are parallel to rising tones. It is the first part of the tone that is consistently aligned with the segmental string, while the second part of the pitch accent seems to be more loosely coordinated with segmental landmarks. The paper makes two main claims. Firstly, alignment in falling accents shows the same asymmetry as that observed in rising accents, namely, that alignment is tighter at the start of the syllable than at the end. Secondly, there is an effect of syllable structure on alignment, since the second pitch target in the bitonal HL accent (the L) aligns earlier in closed syllables than in open syllables.

The aim of Grice, Baumann and Jagdfeld’s paper, “Tonal Association and Derived Nuclear Accents—The Case of Downstepping Contours in German,” is to contribute to the debate on the modeling of intonational accents in German. The study uses phonetic implementation data to test phonological categoriality, arguing for an analysis of downstep in German that involves floating H tones. More generally, the paper argues for a more abstract analysis of intonation, and for the use of floating tones to account for alternations in intonational form. The paper thus contributes to current debates on the modeling of German intonation, as well as to the issue of intonation modeling in general, by providing an account of how surface phenomena (the choice between plain downstep and early peak downstep as a language-specific form of variation in German) can be derived from one underlying structure. In addressing this theoretical issue, the paper also tackles a methodological issue, namely, how tones relate to tonal register and which arguments can be found for tonal modeling based on experimental results.

In particular Grice et al. discuss whether German intonation distinguishes between three (H* L, H+!H*, and H+L*) or two (H* L-, and H+!H*) falling accents. The paper also manipulates information structure variables in order to provide independent evidence for the claim that the falling H+!H* accent is a derived accent, not a basic one. The implementation questions that are touched on by Grice et al. concern both horizontal timing of tonal targets, i.e., alignment of tones relative to the segmental string, and excursion measurements, i.e., the span that relates to the tonal register of a speaker (Ladd, 1996). The alignment measurements are used to confirm a categorical distinction between H* accents and H+ leading tones in falling accents in their recordings. These conclusions are in line with Kügler’s (2007) findings on derived accents in the Upper Saxon variety of German. In order to decide between a H+!H* and a H+L* falling melody, Grice et al. compare neutral with lively speech, assuming that in lively speech a greater pitch range would affect low tones in a different manner than high tones, allowing the categories to be better distinguished. Grice et al.’s conclusion is that German employs H+!H* accents rather
than H+L* accents, a finding which is in line with Grabe (1998), who in a different experimental setting provides evidence for a continuum of downstepped accents.

The general aim of the paper by Yi Xu, “Timing and coordination in tone and intonation—An articulatory-functional perspective,” is to differentiate between timing effects that are based on and constrained by the articulatory mechanism, and timing effects that can be used to express communicative functions. The paper uses the Parallel Encoding and Target Approximation (PENTA) model of speech coding (Xu, 2005) to distinguish between two types of timing effects, obligatory timing on the one hand and informational timing on the other.

Obligatory timing is constrained by the articulatory mechanism. Assuming articulatory targets (i.e., the PENTA model), Xu argues that the minimal movement duration the articulatory process needs to implement a certain articulatory gesture is crucial in determining timing. Tonal articulation is constrained by the ability of our vocal cords to implement transitions from one lexical tone to another. The transition time thus represents the minimum obligatory timing which is needed to implement another tone. This timing then cannot contribute to the functional load of a lexical tone, as it is mandatory. Obligatory timing appears to be aligned within the domain of the syllable. Independent of a preceding tonal target, any movement to a particular tonal target on a syllable starts at the same time, namely, at the beginning of the syllable. Xu argues that the laryngeal timing effects are closely connected with the timing structure of the syllable itself. Any tonal gesture starts when the onset and vowel gesture starts. These facts accord with Prieto’s (this issue) proposal that the initial tone of a bitonal pitch accent is more stably aligned with the syllable than the second tone, independent of its tonal value (high or low).

Informational timing is assumed to represent any timing effect that a speaker actively controls for the purpose of expressing a linguistic function. Given that tonal implementation is constrained by minimal movement duration as well as the timing structure of the syllable, Xu claims that microprosodic variation, i.e., alignment differences, do not contribute to the expression of functional differences (certainly a controversial proposal from the point of view of intonation languages). Informational timing is encoded by means of duration, i.e., changes in the length of segments to express lexical contrast (quantity), or changes in the length of lexical items to express focus. Based on the distinction between obligatory and informational timing, Xu thus provides insights into the very nature of tones and their implementation.

3. Pragmatic aspects of tone and intonation

The articles by Rialland, Schwarz, and Grice et al. address the interface of sentence-level tone – that is, intonation – and semantics/pragmatics. They address questions such as how changes in pitch affect the information structuring of a sentence, or how changes in pitch help to distinguish statements from questions if there is no other linguistic indicator. (See works like Ladd, 1996; Gussenhoven, 2004; Yip, 2002 for recent surveys of these topics.)

The contribution by Rialland, “African lax question prosody: its realization and geographical distribution,” addresses a pragmatic aspect of sentence-level tone, namely the linguistic marking of questions. Cross-linguistically, yes/no questions (polar questions) are most commonly marked by rising intonation (Ultan, 1978; Cruttenden, 1997; Gussenhoven, 2004). Raised intonation reflects uncertainty on the part of the speaker. It is well known that languages differ as to the exact implementation of raised pitch, with higher pitch towards the end of the question seemingly being the prominent pattern (cf. the language sample in Ultan, 1978), as opposed to higher pitch towards the beginning of the question (Siemund, 2001).
Rialland describes a divergent intonation pattern for yes/no questions, one which is characterized by a falling or low pitch contour, breathy termination or a final low tone. She terms this the “lax” question intonation, referring to a lesser degree of tension in the vocal tract during production. Her contribution carefully documents the occurrence of this question intonation pattern in 57 languages out of a database comprising 117 African languages from 3 different African language phyla. (Only Khoisan languages do not show this feature.) The distribution of the “lax” question intonation pattern seems, however, to be restricted to the languages spoken around the equator.

Rialland’s contribution makes three main points. Firstly, the article provides new data, much of it first-hand and unpublished, of relevance both for language-specific and typological studies on question intonation in polar questions. It also motivates further investigation of this aspect of intonation in and beyond the languages cited. However, the data presented in the article suffice to make the second point, namely, that high pitch in polar questions is not a language universal. Thus, the pragmatic explanation for high pitch also cannot be universal: either polar questions do not universally encode uncertainty linguistically, or uncertainty is not cross-linguistically linked to raised pitch. (See Gussenhoven, 2004 for discussion of this point.) Thirdly, Rialland’s contribution suggests that “lax” question intonation is an areal feature which relates the languages of the Sudanic belt (along with other features argued for ongoing research by this author), as it is common to a number of genetically unrelated languages in this region of Africa.

The contributions by Schwarz and by Grice et al. explore the aspects of the interface between tone/intonation and information structure, paying particular attention to the interface between sentence-level tone and focus.

The article by Schwarz, “Tonal focus reflections in Buli and some Gur relatives,” investigates the expression of focus in Buli, a West African Gur language. Buli is a tone language, using changes in pitch to convey lexical and grammatical meaning, and it employs (morpho)syntactic means to mark a constituent as being narrowly focused. In her article Schwarz addresses the question of whether prosodic means are also used to mark the focused constituent in this language. Previous research on the Bantu language Chichewa (e.g., Kanerva, 1990; Downing et al., 2004) and on Chinese (Xu, 1999) has shown that tone languages may indeed use prosodic means to highlight a focused constituent. Instead of pitch accent placement as in Germanic languages, these tone languages use other prosodic means, such as expansion of the overall pitch range (Xu, 1999), tone sandhi due to focus-induced phrase-boundaries (Kanerva, 1990), or manipulation of downdrift (Downing et al., 2004). Schwarz shows, however, that Buli does not make use of comparable prosodic means—or of any other prosodic means, for that matter. Based on exploratory pitch track analysis of question/answer-pairs used to elicit narrow focus, Schwarz shows that there is no evidence for prosodic focus marking in Buli and related Gur languages. These Gur languages thus lack prosodic expression of narrow focus.

Finding that there is no prosodic focus marking in these Gur languages confirms recent research which shows comparable results for an increasing number of other languages, not only African languages of other language phyla (see, e.g., Zerbian, 2007 for the Bantu language Northern Sotho, or Hartmann and Zimmermann, 2007, for the Chadic language Hausa), but also other languages world-wide (see, e.g., Swerts, 2007 for Romanian, Swerts et al., 2002 for Italian, and Kügler and Skopeteas, 2007 for Yucatec Maya). Studies on the well-investigated Germanic languages have shown that pitch accent placement is typically used to mark discourse-new information. Based on such findings, it has been argued in work like Truckenbrodt (1999), Selkirk (2004) and Samek-Lodovici (2005) that focus prominence is a universal. However, typological evidence gathered by studies such as that of Schwarz refutes this claim.
Grice et al.’s article also addresses the influence of information structure on sentence-intonation, as well as the modeling of intonational accent described in the previous section. The second experiment that the authors report on deals with derived nuclear accents in different information structural contexts. It investigates how different focus structures, namely broad, narrow and contrastive focus, are marked in German. Results from a highly controlled, quantitative study show that downsteps can be found more frequently in broad focus contexts whereas upsteps occur more often with contrastive focus in German. However, the study clearly reveals that speakers make individual choices as to their preferred contour. Grice et al. conclude from these results that the downstep relation between two tones is what is crucial in the marking of information structure rather than any particular accent type. Grice et al.’s study encourages research into speaker-dependent variation in the encoding of information structure in order to arrive at a finer-grained model of the prosodic implementation of information structural categories.

4. Conclusion

The papers in this issue, then, make a variety of contributions to our more general understanding of the linguistic use of pitch. First, they bring new data to bear in several areas related to the uses and realization of linguistically meaningful pitch, from segment level interactions between tone and voice to the pragmatic use of tone and intonation at the sentence level. Further, several of the papers confirm the overlap between tonal and intonation languages mentioned in the introduction. For example, the relative salience of tone as a cue to voicing is changing in Kera under influence from non-tonal French. Downstep, a phenomenon that is widespread in tonal languages, plays a role in cuing information structure in the realization of intonation in German. A significant group of African tone languages spoken near the equator have a distinctive yes–no question intonation. Several of the papers challenge widely held typological assumptions about the role of pitch, for example, that pitch prominence (or some other prosody) should be a universal correlate of focus, or that yes–no question intonation should be cued by pitch raising. Others reveal new typological generalizations, for example, the importance of the syllable (especially syllable-initial position) in tonal timing and alignment. It is hoped that both these new generalizations and challenges to old ones provide a stimulus to future research on both tone and intonation.

References