Prosodic Categories: Production, Perception and Comprehension

Book · December 2011

DOI: 10.1007/978-94-007-0137-3

3 authors:

Sonia Frota
University of Lisbon

Gorka Elordieta
Universidad del País Vasco / Euskal Herriko Unibertsitatea

Pilar Prieto
University Pompeu Fabra

Metadata of the book that will be visualized online

BookTitle	Prosodic Categories: Production, Perception and Comprehension	
Book Sub-Title		
Book CopyRight - Year	Springer Science+Business Media B.V. 2011 (This will be the copyright line in the final PDF)	
Series Name	Studies in Natural Language and Linguistic Theory	
Book Editor	Family Name	**Frota**
	Particle	
	Given Name	**Sónia**
	Suffix	
	Email	sonia.frota@mail.telepac.pt
	Organization	Universidade de Lisboa
	Division	Dept. Linguistica Laboratorio de Fonetica da FLUL Alameda de Universidade
	Address	1600-214, Lisboa, Portugal
Book Editor	Family Name	**Elordieta**
	Particle	
	Given Name	**Gorka**
	Suffix	
	Email	gorka.elordieta@ehu.es
	Organization	Euskal Herriko Unibertsitatea, Letren Fakultatea Hizkuntzalaritza eta Euskal Ikasketak
	Division	
	Address	Unibersitatearen ibilbidea 5, 01006, Vitoria-Gasteiz, Spain
Book Editor	Family Name	**Prieto**
	Particle	
	Given Name	**Pilar**
	Suffix	
	Email	pilar.prieto@upf.edu
	Organization	ICREA-UPF, Campus de la Communicació-Poblenou
	Division	Departament de Traducció i Ciències del
	Address	Carrer Roc Boronat 138, 08018, Despatx 53.710, Barcelona, Spain

PROSODIC CATEGORIES: PRODUCTION, PERCEPTION AND COMPREHENSION

Studies in Natural Language and Linguistic Theory

VOLUME 82

For further volumes:
http://www.springer.com/series/6559

PROSODIC CATEGORIES: PRODUCTION, PERCEPTION AND COMPREHENSION

by

Sónia Frota
Universidade de Lisboa, Portugal

Gorka Elordieta
Euska Herriko Unibertsitatea, Spain

Pilar Prieto
Institut Català de Recerca i Estudis Avançats & Universitat Pompeu Fabra, Barcelona, Spain

Springer

Editors

Sónia Frota
Universidade de Lisboa
Dept. Linguistica
Laboratorio de Fonetica da FLUL
Alameda de Universidade
1600-214 Lisboa
Portugal
sonia.frota@mail.telepac.pt

Gorka Elordieta
Euskal Herriko Unibertsitatea
Letren Fakultatea
Hizkuntzalaritza eta Euskal
Ikasketak
Unibersitatearen ibilbidea 5
01006 Vitoria-Gasteiz
Spain
gorka.elordieta@ehu.es

Pilar Prieto
ICREA-UPF
Campus de la Communicació-Poblenou
Departament de Traducció i
Ciències del
Carrer Roc Boronat 138
08018 Barcelona
Despatx 53.710
Spain
pilar.prieto@upf.edu

ISSN 0924-4670
ISBN 978-94-007-0136-6 e-ISBN 978-94-007-0137-3
DOI 10.1007/978-94-007-0137-3
Springer Dordrecht Heidelberg London New York

Library of Congress Control Number: Applied for PCN

Printed on acid-free paper

Springer is part of Springer Science + Business Media (www.springer.com)

Contents

Contributors

Jüri Allik Institute of Psychology, University of Tartu, Tartu, Estonia, juri.allik@ut.ee

Aoju Chen Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands, aoju.chen@mpi.nl

Yiya Chen Leiden University, Leiden, The Netherlands, yiya.chen@hum.leidenuniv.nl

Laura J. Downing ZAS, Berlin, Germany, downing@zas.gwz-berlin.de

Gorka Elordieta Euskal Herriko Unibertsitatea-Universitat de les Illes Balears, Palma, Spain, gorka.elordieta@ehu.es

Anouschka Foltz Department of Linguistics, Ohio State University, Columbus, OH, USA, anouschka@ling.ohio-state.edu

Sónia Frota Laboratário de Fonética & Lisbon Baby Lab (CLUL/FLUL), Universidade de Lisboa, Lisboa, Portugal, sfrota@fl.ul.pt

Martine Grice Department of IfL-Phonetics, University of Cologne, Cologne, Germany, martine.grice@uni-koeln.de

Larry M. Hyman Department of Linguistics, University of California, CA 94720-2650, Berkeley, USA, hyman@berkeley.edu

Mariapaola D'Imperio Laboratoire Parole et Langage, UMR 6057 CNRS, Université de Provence (Aix-Marseille I), Aix-en-Provence, France, mariapaola.dimperio@lpl-aix.fr

Kiwako Ito Department of Linguistics, Ohio State University, Columbus, OH, USA, ito@ling.ohio-state.edu

Kyunghee Kim Department of IfL-Phonetics, University of Cologne, Cologne, Germany, kimk@smail.uni-koeln.de

Karsten A. Koch Research Fellow, Zentrum für Allgemeine Sprachwissenschaft, Berlin, Germany, karstenak@hotmail.com

Aditi Lahiri Faculty of Linguistics, Philology and Phonetics, University of Oxford, Oxford, UK, aditi.lahiri@ling-phil.ox.ac.uk

Pärtel Lippus Institute of Estonian and General Linguistics, University of Tartu, Tartu, Estonia, partel.lippus@ut.ee

Kristine Maday Department of Linguistics, Ohio State University, Columbus, OH, USA, kristine.maday@gmail.com

Kemmonye C. Monaka The Department of English, University of Botswana, Private Bag 00703, Gaborone, Botswana, MONAKA@mopipi.ub.bw

Karl Pajusalu Institute of Estonian and General Linguistics, University of Tartu, Tartu, Estonia, karl.pajusalu@ut.ee

Caterina Petrone Zentrum für Allgemeine Sprachwissenschaft, Berlin, Germany, petrone@zas.gwz-berlin.de

Pilar Prieto Department de Traducció i Ciències del Llenguatge, Institut Català de Recerca i Estudis Avançats-Universitat Pompeu Fabra, Barcelona, Spain, pilar.prieto@upf.edu

Shari R. Speer Department of Linguistics, Ohio State University, Columbus, OH, USA, speer@ling.ohio-state.edu

Michelina Savino Department of Psychology, University of Bari, Bari, Italy, m.savino@psico.uniba.it

Linda Wheeldon School of Psychology, University of Birmingham, Birmingham, UK, l.r.wheeldon@bham.ac.uk

Introduction

Sónia Frota, Gorka Elordieta, and Pilar Prieto

As the title indicates, *Prosodic Categories: Production, Perception and Comprehension* addresses the central question of the role played by prosody in language grammar and language processing. The eleven chapters of this book were developed from presentations to the *Third Tone and Intonation in Europe Conference* (TIE3), hosted by the Universidade de Lisboa, Portugal, in September 2008, and all of them deal with different aspects of the definition, implementation and processing of prosodic categories. They present novel contributions to the understanding of key issues in prosodic theory, such as prosodic phrasing in production and comprehension, the relationship between intonation and pragmatics in speech production, speech perception and comprehension, the development of prosodic categories that convey specific pragmatic meanings, the characterization of the prosody of sentence modality, the role of pitch in quantity-based sound systems, the phonology of consonant conditioned tone depression across languages, and the encoding of intonational contrasts both in intonational and in tonal languages.

Exploring the intersection of phonology, phonetics and psycholinguistics, most of the chapters draw on empirical approaches to prosodic patterns in language: in particular, production, perception and comprehension experiments which include the prepared speech paradigm, the on-line speech production paradigm, conversational style and picture-naming production tasks, eye-tracking experiments using the real-world object manipulation paradigm, identification, discrimination and semantic scaling tasks, as well as perceptual experiments resorting to the gating paradigm. The production, perception and comprehension of prosodic categories is discussed in a wide array of languages (Swedish, Norwegian, Dutch, English, Bari Italian, Neapolitan Italian, Bengali, Estonian, Korean, Shanghai Chinese, Zulu, Shekgalagari, and Nɬeʔkepmxcin), some of them underrepresented in the literature and others

S. Frota (✉)
Laboratório de Fonética & Lisbon Baby Lab (CLUL/FLUL), Universidade de Lisboa, Lisboa, Portugal
e-mail: sfrota@fl.ul.pt

S. Frota et al. (eds.), *Prosodic Categories: Production, Perception and Comprehension*, Studies in Natural Language and Linguistic Theory 82,
DOI 10.1007/978-94-007-0137-3_1, © Springer Science+Business Media B.V. 2011

described for the first time with regard to these topics (like Nɬeʔkepmxcin and Shekgalagari).

Aditi Lahiri and Linda Wheeldon's chapter investigates the prosodic grouping of phonological clitics and compound words, based on synchronic and diachronic data, as well as psycholinguistic evidence. They ask two key questions: do languages exhibit a preferred grouping of lexical words into larger prosodic constituents based on rhythmic principles? If so, does this prosodic grouping play a role in language production planning, that is in the processing involved in planning to produce speech? The authors entertain the hypothesis that, at least in a subset of languages, the natural grouping is trochaic, namely function words cliticize leftwards and compound words are left-headed. They set out to convincingly show that in Swedish, Norwegian, Dutch, English and Bengali leftwards attachment is indeed the natural prosodic grouping. Evidence comes from language change data showing the encliticization of the definite article in Scandinavian and the encliticization of auxiliaries in Germanic and Bengali. Additional evidence is provided by enclitization in Dutch, and unstressed words in English are also argued to show left attachment as their default pattern of prosodic phrasing. In short, trochaic grouping is claimed to be the preferred pattern for this set of languages, thus yielding phonological words that include clitics before their right edge and compound-like units that group together two phonological words.

The authors further show that research on the planning and articulation of speech allows the nature and size of the units the speaker uses in speech planning to be determined. They present experimental evidence, both from the prepared speech paradigm and the on-line speech production paradigm, that strongly suggests that the relevant unit for planning is the phonological word, and that the direction of attachment during clicitization (in Dutch) is leftwards. Specifically, on the basis of prepared speech production studies, they concluded that speech onset latency was a function of the number of phonological words in the utterance, and not of the number of lexical items, and that two word compounds are treated as one phonological word. The on-line speech production studies provided a clear indication that encliticization is the resultant prosodic grouping, as onset latencies to sentences that start with a word plus a clitic were slower than to sentences without a clitic. Taken together, the findings reported strongly support the hypothesis of trochaic grouping. However, as Lahiri and Wheeldon remark, this form of grouping is not universal: crucially, Romance languages have been shown to be by and large proclitic, favoring rightwards attachment of clitics and exhibiting compound-like structures which are right-headed (PeperKamp 1997; Vigário 2003; Hualde 2006/2007). Moreover, whereas Germanic languages extend the unstressed stretch on the right-hand side of phonological words by attracting unstressed elements, Romance languages tend to lose segments or syllables after the word stress (either by deletion or resyllabification – Harris 1988; Vigário 2003; Wheeler 2005). The reasons behind the difference in preferred prosodic grouping across languages (and related phenomena) constitute an important area of research with implications for different fields, such as language typology, language variation and change and language acquisition.

Examining prosodic groupings at a higher level, Foltz, Maday and Ito's chapter is devoted to the production and processing of sentences with structural ambiguities. In their paper, these authors investigate the patterns of prosodic phrasing and semantic interpretation of sentences with two nouns in an NP followed by a relative clause, where the relative clause is sufficiently ambiguous to have either N1 or N2 as antecedents of the empty subject. An example of a sentence of this type would be *The brother of the bridegroom who swims was last seen on Friday night*. The investigation aimed at discovering the role of constituent length in determining prosodic boundaries as well as their strength. By varying the length of N1 (*the brother*, in the sample sentence above) and the relative clause, Foltz, Maday and Ito want to test the hypothesis that long constituents tend to be set off in their own prosodic phrases more often than short constituents (Carlson et al. 2001; Clifton et al. 2002; Fodor 1998, 2002; Watson and Gibson 2005). The other objective of Foltz, Maday and Ito is to see whether a full semantic processing and interpretation of an entire sentence affects prosodic boundary placement in a different way from a partial or online semantic processing. To this end, the authors carried out two production and perception experiments. In Experiment 1, subjects had to read and pronounce the target sentences (with filler sentences) on the fly, as they appeared on the screen, and then respond to a comprehension questionnaire by choosing between N1 and N2 as the antecedent of the subject of the relative clause (i.e., choosing between low and high attachment of the relative clause). In Experiment 2, the comprehension task was carried out first, and then the production task, after the subjects had read and processed the sentences semantically and assigned an antecedent for the subject in the relative clause. A comparison of the results of both experiments would be telling as to whether constituent length modulates prosody in all reading circumstances or only when the global sentential structure and its meanings are fully established by the reader/speaker.

The results for Experiment 1 show that the length of N1 determines prosodic boundary insertion, in the sense that the longer N1 is the greater the likelihood for prosodic boundary placement after it, but the results also show that relative clause length did not affect prosodic boundary placement and strength. Thus, these results do not support the hypothesis that long constituents are more likely to be segmented in an independent prosodic phrase (ip or IP). In the sentence comprehension task following the production task, subjects did not match their choices for antecedents with their production patterns, a result which is surprising in the light of previous work that had reported a closer match between prosodic boundary location and strength and attachment preferences (Carlson et al. 2001; Clifton et al. 2002). Experiment 2 shows that constituent length does not influence attachment preference as strongly as previously assumed, and that the preferred attachment option (low attachment) does not elicit stronger boundaries after N1 as often as expected. Unlike Experiment 1, however, Experiment 2 suggests that the length of the relative clause combines with the length of N1 to affect the placement and strength of prosodic boundaries. That is, the results of Experiment 2 show that a more global consideration of the whole sentence is taking place when subjects can take time to process and comprehend the sentence before reading it aloud. Foltz, Maday and Ito explain the differences between the results of the two experiments in terms of the limits set by the eye-voice span of reading on the fly (as in Experiment 1), which does not allow calculating the length of the relative clause, opposed to the more generous time a subject has when reading a sentence without the pressure to utter it immediately (as in Experiment 2). In Experiment 1, the subjects would be guided by the presence of the syntactic boundary corresponding to the relative clause, more than by its length. In Experiment 2, on the other hand, the global knowledge of the syntactic and semantic structure of the sentence seemed to allow for a better rhythmic adjustment of the utterance.

The comparison between the two contexts for sentence production (reading on-line and after sentence comprehension) established by Foltz, Maday and Ito constitutes a novelty in studies of prosodic boundary insertion and their (mis) match with attachment preferences. The results and main conclusions of this work help advance our knowledge of the matching between sentence processing or comprehension and sentence production, as regards structurally ambiguous sentences. In particular, the results raise interesting questions as to how silent reading operates and how silent reading may reveal a lack of uniformity in implicit prosody that was not assumed before (cf. Fodor 1998, 2002; Jun 2003).

The next four chapters are all concerned with various aspects of the prosody of focus in four different languages. Kiwako Ito and Shari R. Speer's chapter has the goal of investigating the effect of the presence of a L + H * pitch accent on pre-nominal adjectives in English. Even though it is well-known that the presence of a salient/focal pitch accent has the role of singling out a specific item from a larger set, what is less known is whether the semantics of the accented adjectives influences the activation of this contrastive interpretation. With this aim in mind, Ito and Speer conducted a pair of eye-tracking experiments designed to compare the effect of the presence of a prominent pitch accent on pre-nominal intersective color adjectives (Color experiment) and subsective size adjectives (Size experiment). Importantly, since subsective adjectives such as *big* or *old* require a relative interpretation, the hypothesis is that these types of adjectives may automatically evoke a notion of contrast (e.g., use of 'big' in 'Give me a big cup', which implies the presence of smaller cups). Contrastive interpretation was tested using the well-known real-world object manipulation paradigm, where participants followed pre-recorded instructions to decorate holiday trees. Eye fixation results revealed that in both experiments (the Color and the Size experiments) the presence of a prominent pitch accent (L + H*) on the adjective facilitated the more rapid detection of the contrastive target (e.g., Hang a red/medium star. - Next, hang a *YELLOW/LARGE* star.). When L + H* was infelicitously used in non-contrastive sequences (e.g., Hang a red/ medium tree. Next, hang a *YELLOW/LARGE* ball.), results indicated no inherent semantic advantage for the subsective adjectives, given that fixation times were not longer in the non-focal accent condition with the size adjectives than with the color adjectives. Moreover, contrary to the authors' prediction, the likelihood to fixate on the contrastive competitor was generally much higher for the Color than for the Size experiment. The results thus demonstrate that the presence of a focal prominence is evaluated online against the discourse context. When L + H* is used felicitously on the adjective, it facilitates the detection of the target, regardless of the semantics of the adjective. By contrast, if L + H* is used in an infelicitous way, it results in a slower detection of the correct target rather than as an increase in the fixations to the incorrect contrastive referents.

Importantly, while Ito & Speer's results confirm that the presence of a focal L + H* pitch accent on pre-nominal modifiers activates a contrastive interpretation, they also found that the online comprehension of contrastive meanings is modulated by the discourse and visual context. Crucially, the unexpected bias toward a contrastive interpretation in the Color experiment uncovered a striking effect of the visual and discourse context in utterance interpretation rather than an inherent difference in adjective semantics. The authors convincingly attribute these differences to the salience/ease of the visual contrast that characterizes the display of the Color experiment as compared to the one of the Size experiment. One important methodological lesson to be learnt from these results, as the authors rightfully point out, is that experimental paradigms need to be controlled for easiness in referential comprehension within the visual field.

Aoju Chen's chapter is concerned with the phonological focus-marking in child language in Dutch. The information structure category focus, defined as the constituent that expresses new information in a sentence, exhibits specific phonological cues that are essential to focus-marking in adult Dutch: namely, information structure constrains both the placement of pitch accents and pitch accent choices in focused and unfocused positions. Children's ability to use phonological cues to mark non-contrastive narrow focus early on is inspected, in spontaneous (two-year-olds) and elicited speech (four- to eight-year-olds). On the methodological side, Chen's study puts forward an ingenious experimental set-up for data elicitation, that allows the collection of strictly comparable SVO declaratives with focused and unfocused NPs both in sentence-initial and sentence-final position. This is achieved by means of a picture-matching game where the production of trigger utterances is controlled for intonation pattern. The intonational analysis of child language was based on the Transcription of Dutch Intonation system (ToDI), which although designed to account for adult intonation has proven successful in dealing with child speech. The results have shown that children, two-year-olds included, use phonological means to mark focus, but not in an adult-like way. The use of accent placement and accent type to mark focus is acquired in a gradual fashion: two-year-olds do not use accent placement, but make a difference between non-downstepped accent types, used in focused words, and downstepped accents, used in unfocused contexts; four- to five-year-olds are already adult-like in using accent placement to realise the focused/unfocused contrast, but, unlike adults, show no preference for the H*L accent over the other accent types for focus in sentence-final position; seven- to eight-year-olds are largely adult-like in the phonological marking of focus.

The developmental path to phonological focus-marking proposed thus suggests a first step where focused/unfocused is generally equated to phonetically strong/weak, followed by the acquisition of the relationship between accent placement (accented/deaccented) and information structure, and finally of the relationship between accent type (first in sentence-initial position and later in sentence-final position) and information structure. As noted by the author, further analysis on younger children's utterances with final focus is called for to verify whether the phonetic strength distinction holds in this condition. Moreover, we would add, the prosodic phrasing of two-word utterances at this early stage needs to be carefully considered, as it may be the case that accent placement is not used to mark focus because each word forms its own phrase (as in single word utterances, or in successive single word utterances – Hallé et al. 1991; Behrens and Gut 2005; Frota 2010b). The proposed developmental path raises interesting questions for languages unlike Dutch, such as French (where pitch accent shape seems to play no role in focus marking), many other Romance languages (where pitch accent type matters but the Germanic accented/deaccented contrast does not exist in the intonation grammar), or Mandarim Chinese (where focus is marked by pitch range variation and duration). The effects of language-specific input on the path of development are a challenging topic for future cross-linguistic research on the interaction between prosodic categories and information structure in acquisition.

The contribution by Karsten Koch also deals with the interaction of intonation and information structure. Specifically, Koch provides a detailed phonetic analysis of prominence cues to focus and given information in Nłeʔkepmxcin (Thompson River Salish). This is the first such study in any Salish language, thus providing new data as well as relevant insights from a yet largely unstudied endangered language. Based on informal and impressionist observations from the literature, the author sets out to test the hypothesis that Nłeʔkepmxcin, although a stress language, does not mark the focus/givenness distinction by means of pitch accents or any kind of additional prominence/reduction of prominence in comparison with neutral, wide focus. The focus types under analysis are subject and object narrow focus, both non-contrastive and contrastive, and wide focus or focus-neutral. The examination of the most common acoustic prosodic correlates of focus and givenness, as reported in the literature on stress languages – namely, peak height, peak timing, local pitch range, peak intensity and accented vowel duration –, shows the absence of pitch cues in the marking of focus and givenness (the results of intensity and duration being inconclusive). This result, as argued by the author, makes Thompson Salish a typologically unusual stress-accent language, and has implications for the putative universality of constraints like STRESS-FOCUS and DESTRESS-GIVEN (e.g. Féry and Samek-Lodovici 2006), within stress languages.

The study undertaken by Koch, similarly to the work by Chen, points to the importance of cross-linguistic experimental research to establish the prosodic categories involved in the marking of information structure. In the case of Koch's study, the absence of acoustic pitch cues to prominence that are common in many stress languages is undoubtedly a relevant finding, although a phonological analysis of the Thompson Salish intonation system is not within the scope of this study. Such a finding also highlights the need for a global grammatical approach to focus marking. Thompson Salish, similarly to other languages where prominence cues are not used for focus marking, such as Wollof (Rialland and Robert 2001), marks focus by morpho-syntactic means, and thus it could be argued that the use of stress and accent features is analogous in grammatical function to the use of morphological and syntactic focus marking (Frota 2000, 2002). Under this view, prosodic categories need not be universally involved in focus marking, and their role in cuing information structure may have been overestimated by the study of Indo-European stress languages, as Koch duly suggests.

The chapter contributed by Kyunghee Kim investigates the influence of a number of prosodic and non-prosodic factors on the alignment patterns of the peak of the Accentual Phrase (AP) in Korean. In this language, the AP is one of the main units of prosodic analysis and it is demarcated by the tonal pattern THLH (T = L or H). Thus the AP contours can contain a peak which corresponds to an initial phonological H tone associated with the second syllable of an AP. The peak is phonetically realized either in the second or third syllables of the AP, and little is known about which are the factors that condition the alignment of this peak (see previous work on Korean intonation by Jun 2005). Two production experiments were conducted which consisted of casual style conversations with questions and elicited target answers. Experiment 1 aimed at investigating the potential effect of the number of phonological words in the AP, sentence length (presence or absence of a preceding AP), and focus type (narrow focus vs. broad focus) on peak placement. Results revealed that the realisation of narrow focus depends on AP length. In the short two-word AP, narrow focus is realised by earlier peak alignment. Number of phonological words in the AP has a significant effect on peak alignment, which is placed earlier in one-word APs than in two-word APs. Unexpectedly, the peak in the one-word AP was located systematically in the third syllable, as in two-word APs, indicating that the presence of a morpheme boundary is likely to affect the peak alignment. To test this hypothesis, Experiment 2 was carried out. The goal of this experiment was to test whether the alignment of the accentual peak was affected by the presence of an upcoming morpheme boundary and the presence/absence of semantic content in the following morpheme. The results indeed showed that accentual peak alignment is significantly affected by the presence of a morpheme boundary and by the semantic content of the following morpheme.

One of the most interesting results of the chapter by Kyunghee Kim is the fact that the alignment pattern of the accentual peak is systematically affected the location of a morpheme boundary. The location of the AP peak in Korean is confined to the AP initial morpheme, and the peak is aligned later as the morpheme becomes longer. Thus as the author argues, the H tone is associated with an edge of a morpheme. Yet this constraint is overridden by the semantic importance of the following morpheme. Even though previous work on tonal alignment patterns in European languages has highlighted the importance of prosodic factors on peak location, the results of this chapter demonstrate that in other languages the location of *f0* peaks may directly depend on the presence of upcoming morpheme boundaries. Recent work by Prieto et al. (2010) has reported a similar phenomenon in two Romance languages, namely Spanish and Catalan. In these two languages, the presence of word boundaries also affects prenuclear peak location, and the target position of the peak can even be used by Catalan and Spanish for word identification. These crosslinguistic findings stress the importance of carrying out typological work and encourage promising work in this area.

The following two chapters report on research on the perception of intonation categories and their relevance to meaning distinctions. Michelina Savino and Martine Grice conduct two perception experiments to investigate the role of pitch height variation in distinguishing between two different questions types in the Bari variety of Italian, namely yes-no information seeking question (Queries) and questions that challenge assumed given information (Objects). Previous work on the intonational marking of questions in Bari Italian has shown pragmatic differences that are implemented in a gradient fashion may exhibit a discrete intonational marking, as in the case of the Query-Check dimension where different degrees of speaker's confidence are signalled by means of contrasting pitch accents (L + H*, H* + L, H + L*). The difference between Objects and the other questions types is also to be found in the accented syllable. Indeed, the nuclear pitch accent is the domain for the intonational marking of questions in Bari Italian (and not the boundary tone). However, Objects show what appears to be the same pitch accent as Queries (L + H*), but with a higher peak. The question thus arises whether the pragmatic distinction between Objects and Queries is signalled gradiently by intonation, or whether the peak height difference is phonological. Both an identification task and a discrimination task were carried out to examine whether the peak height difference is perceived categorically. Along the lines of other studies on pitch height differences in several languages (e.g. Chen 2003; Falé and Hub Faria 2006; Borràs-Comes et al. 2010), listeners responses in the semantically motivated identification task, together with the obtained pattern of Reaction Time measurements, clearly show that they are able to make a categorical interpretation of utterances as Query or Object on the basis of peak height variation only; however, and again in accord with many other studies, the results of the purely psychoacoustic discrimination task show that listeners are unable to discriminate between pairs of stimuli.

The results are interpreted in two relevant ways. First, the success obtained in the semantically motivated task, which is truly a linguistic task where subjects must access linguistic knowledge on the categories available in the language and the way they are realized, points to the presence of two categories and therefore to the need to represent pitch height in the intonational phonology of Bari Italian. Second, the failure in the discrimination task, which is a psychoacoustic task that did not involve accessing linguistic knowledge, strongly suggests that this kind of task is not suitable to investigate intonational meaning contrasts, which have a semantic/pragmatic nature (see, as well, Chen 2003; Frota 2010a). The authors point to listener specific competence and acoustic memory restrictions as possible key factors affecting discrimination performance. In other studies, it has been argued that the problem may reside in the non-linguistic nature of the task, as semantically motivated discrimination tasks have been shown to provide different results (Schneider et al. 2009; Frota 2010a). The present chapter does offer an important contribution to the discussion about the approaches and methods to define prosodic categories. The chapter also questions the phonological nature of the intonational contrast investigated. In this study, as the authors rightfully mention, peak height is strictly equivalent to pitch range, and thus the results obtained do not directly inform the kind of phonological analysis that may best account for the pitch height categorical distinction, calling for future investigation.

The chapter written by Caterina Petrone and Mariapaola D'Imperio deals with the potential difference found between the prenuclear contours of Neapolitan Italian narrow focus statements and yes/no questions. As it is well-known from previous work by Mariapaola d'Imperio and colleagues (D'Imperio 2000; D'Imperio and House 1997), these two sentence types are distinguished by nuclear pitch accent alignment, namely early alignment (L + H *) is found in narrow focus statements and late alignment (L* + H) is found in yes/no questions. The chapter investigates two related questions: (a) the potential relevance of a tone which appears to be inserted at the right edge of the Accentual Phrase in the prenuclear contour (H in questions and L in statements), testing whether Neapolitan listeners are able to identify the contrast between questions and statements in the prenuclear region based on the edge tone difference; (b) the linguistic and paralinguistic meaning differences conveyed by this prenuclear edge tone. To study these questions, two experiments were run with gated stimuli, one with a forced-choice identification task and the other using a set of five semantic differential tasks. In the first experiment (Experiment 1), nine Neapolitan listeners heard three gates of the two sentence types for three separate sentences (with a control that contained the whole sentence) and they were asked to label each stimulus as either a question or a statement. If question/statement identification depended solely on the availability of nuclear accent information, listeners would not be capable to identify the target sentence type in early gates. Yet the results showed that the presence of prenuclear accent information was important: mean 'question' score for question base stimuli was above chance (67%), while it was around 37% for statement base stimuli. Moreover, when the prenuclear edge tone was present in the gate, question scores decreased for statement base stimuli (20%), suggesting that the presence of this tone plays an important role in sentence type identification. The second experiment (Experiment 2) explored the potential contribution of the scaling of the target prenuclear edge tones L and H on the linguistic or paralinguistic meaning of this part of the contour. A set of five semantic differential tasks was run with five semantic scales that were selected on the basis of hypotheses about the linguistic and paralinguistic properties of the two accentual phrase tones, namely, 'commitment' (belief about the compatibility of the speaker's beliefs with those of the listener), 'potency' (or certainty or uncertainty about the content of his/her message), 'activity' (speakers' emotional involvement), 'evaluation' (speakers' sociability), and 'submission' (speakers' degree of authoritativeness). Nine Neapolitan listeners heard a set of gated utterances which contained a continuum in scaling between the L and the H prenuclear edge tones. The results revealed that only two out of the five scales varied with the scaling changes. Specifically, low edge values conveyed a higher degree of 'certainty', which progressively decreased as the edge height increased. Also, the mean involvement score progressively decreased as the height of the tone increased. The results of the two experiments pose a series of interesting questions to the study of prosody and meaning. While differences in the perception of the two modalities are demonstrated to be at work already in very early portions of the utterance and crucially at the point where the edge tone is available (see results of Experiment 1), differences in scaling of the edge tone significantly affect other listeners' semantic judgments.

This chapter represents a contribution to the still understudied topic of the relevance of prenuclear contours to meaning, as recent work on intonational phonology has mainly focused on the role of nuclear configurations. The results of Experiment 1 clearly challenge the idea that the nuclear configuration is the only cue that is relevant for the question-statement distinction. Tune meaning is clearly the result of the interaction between prenuclear and nuclear *f0* contours, and this issue calls for further investigation, as the authors point out. The results of Experiment 2 also provide with an interesting groundword for starting the investigation of the relationship between the linguistic and paralinguistic semantic weight carried by prenuclear and nuclear contours in tune interpretation.

The chapter contributed by Pärtel Lippus, Karl Pajusalu and Jüri Allik deals with one of the central features of Estonian word prosody, namely, the three-way distinction in quantity (short or Q1, long or Q2, and overlong or Q3, which can be implemented by both vowels and consonants). While the primary cue for the three-way quantity distinction is the temporal pattern of the disyllabic foot, pitch patterns also seem to play an important role in the identification of the three-way quantity system. While in short and long categories the pitch falls at the end of the stressed syllable, in the overlong category the pitch falls in the first half of the stressed syllable. Previous research has shown that a conflicting combination of pitch and temporal cues can significantly affect target word identification. The main aim of the chapter is to investigate the role of pitch contour changes in the perception of the long vs. overlong quantity distinction. The authors performed two identification experiments in which they manipulated different properties of the pitch patterns of a Q2 word without changing its duration. In Experiment 1, the locus of the fall was always in the middle of the accented syllable and the duration of the fall was varied in five steps. By contrast, in Experiment 2, the start of the fall was varied by five 20 ms increments and the pitch always fell during 50 ms (about 1/3 of the target vowel duration). In both experiments, listeners had to decide about the meaning of the words: ‘send!’ in case of Q2 and ‘to get’ in case of Q3. The results of Experiment 1 demonstrate that the duration of the fall is important for Q3 recognition, as the fall cannot be too short. The results of Experiment 2 show that the locus of the fall is crucial for Q3 perception. The pitch contour that triggers Q3 perception falls in the middle of the target vowel; yet if the pitch fall is too early or late during the vowel, Q2 is perceived. Pitch range is also found to be an important cue to quantity distinctions.

The chapter by Pärtel Lippus and collaborators represents an important contribution to our knowledge about the combined role of tonal and duration prosodic cues in the conveyance of lexical meaning in a language with a three-way quantity distinction. As the authors point out, the results of both experiments showed that several tonal features are essential for the perception of Q3, namely a significant locus of the falling *f0* pattern, pitch range of the fall and optimal length of the pitch movements. The potential interaction between these acoustic parameters is a challenging task for the future and for our understanding of the relationship between duration and *f0* parameters in the perception of speech.

The two last chapters in this volume discuss the phonology and phonetics of the uses of *f0* in tonal languages. Yiya Chen and Laura J. Downing establish a phonetic and phonological comparison of tone depressor consonants in two typologically unrelated languages, Shanghai Chinese and Zulu. Their objective is to rebate a previous proposal by Jessen and Roux (2002) that the depressor consonants in Xhosa (another Nguni language closely related to Zulu) and Shanghai Chinese can be characterized by the feature [slack voice], implemented phonetically the same way in the two languages, and that *f0* lowering occurs to compensate for absence of phonetic voicing of the depressor consonants. The data of Shanghai Chinese comes from previous work by Chen (2007), in which the effect of consonantal laryngeal features (aspirated, unaspirated and depressor) in different tone combinations across syllables was investigated. The data of Zulu comes from an experiment specifically designed by Chen and Downing to test the phonetic effect of depressor consonants, following Chen’s (2007) methodology. The results of the experiment for Zulu show that the *f0* lowering effect of depressor consonants is different from that of depressor consonants in Shanghai Chinese; whereas in Zulu the *f0* level maintains low throughout the target syllable, in Shanghai Chinese it wanes much faster during the target syllable. Moreover, in Zulu *f0* lowering applies word-initially and word-medially, whereas in Shanghai Chinese there is less *f0* lowering word-medially. These results contradict Jessen and Roux’s (2002) previous claim that depressor consonants behave the same way in the two languages. Chen and Downing also show that Zulu implosives, which are fully voiced, do not lower *f0* as depressor consonants do. This fact argues against a possible explanation of the *f0* lowering effect of depressor consonants as a compensation for the loss of voicing of these consonants, as suggested by Jessen and Roux (2002). Chen and Downing propose an alternative account for the differences in *f0* lowering between Shanghai Chinese and Zulu. Like Jessen and Roux (2002), they assume a feature [slack voice] for depressor consonants in the two languages, but they argue that the different phonetic effects this feature has in Shanghai Chinese and Zulu can be explained by phonological differences between the two languages. On the one hand, there are differences between Shanghai Chinese and Zulu regarding the domains of tonal specification: in Shanghai Chinese the domain is the syllable, whereas in Zulu it is the word. This would explain why in Shanghai Chinese the *f0* lowering effect wanes faster than in Zulu. On the other hand, in Shanghai Chinese, the word-medial underlying tonal specifications of each syllable are lost by tonal sandhi processes, including the depressor register, so the different behaviour of depressor consonants word-medially could be explained as a way to compensate for the loss of tonal specifications, that is, as a way to maintain certain phonological specifications. In Zulu, there is no specific depressor register and there are no tonal sandhi processes word-medially that obliterate underlying tonal information, so there is no need to compensate or maintain anything.

Chen and Downing's account of the differences in the effects of *f0* lowering between two tonal languages such as Shanghai Chinese and Zulu finds support in the idea that a single phonological feature can have different phonetic implementations in different languages (cf. Keating 1988; Kenstowicz 1994), and that these differences in phonetic implementation are governed or controlled by the phonology of each language (cf. Kingston and Diehl 1994). For the case at hand, Chen and Downing conclude that the phonological feature [slack voice] is present in Shanghai Chinese and Zulu but that the different phonetic realization of this feature is governed by (higher-order) phonological considerations. Hence, Chen and Downing's chapter contributes to our understanding of the phonological and phonetic properties of depressor consonants in two genetically unrelated tonal languages such as Shanghai Chinese and Zulu, refuting earlier proposals on the issue, as well as to our understanding of the phonology-phonetics interface in general.

In the final chapter, Larry M. Hyman and Kemmonye C. Monaka address the different non-tonal strategies employed by tonal languages to convey grammatical meanings conveyed intonationally in non-tonal languages, such as sentence type. The central issue at stake is that tonal languages make use of *f0* for establishing lexical and grammatical contrasts, so introducing additional tonal events at the phrasal or sentence level may give rise to conflicts with the tonal information already present at the word-level. As Hyman and Monaka point out, word-level tones may adopt three different types or degrees of receptiveness towards phrase- or utterance-level tones: accommodation or coexistence, submission or surrender, and avoidance or blockage. Shekgalagari, a Bantu tonal language, represents the choice of accommodation, according to Hyman and Monaka. In this language, information of sentence type is signalled intonationally for certain sentence types, such as declarative sentences and citation forms, through the insertion of a L% boundary tone that is phonologically associated to the second mora of the lengthened penultimate syllable in phrase-final position. But for other sentence types, non-intonational cues are used. In ideophones, the sentence-final vowel is devoiced. In paused lists, the final syllable of each member of the list is lengthened. And in yes-no questions, wh-questions, imperatives, exhortatives, vocatives, exclamatives and monosyllables no segmental or suprasegmental cues are used. The marking of sentence types in Shekgalagari is interesting for a general theory of grammar, as it shows that the unmarked, general way of marking the majority of sentence types is through the absence of overt cues. That is, it could be concluded that in this language the unmarked cue of sentence type is the absence of intonation. Intimately related to this point, it is worth mentioning that another noteworthy aspect of the Shekgalagari system of marking sentence type is that it associates a phonologically marked cue (overt intonation and penultimate lengthening) to a pragmatically unmarked construction such as a declarative sentence.

Hyman and Monaka also show that when a sentence in Shekgalagari contains more than one of the above mentioned syntactic structures (e.g., when a wh-question ends in an ideophone), some cues override others, and thus the authors describe a hierarchy of sentence types or sentence-type cues. Interestingly, the phonologically unmarked sentence types (yes-no and wh-questions) dominate the marked sentence types (such as ideophones, lists and statements). Additionally, Hyman and Monaka reveal the existence of what they call emphatic declaratives, which are analyzed as abstract declaratives created out of the interrogative or exhortative sentences. Emphatic declaratives are marked by the same cues as regular declaratives or statements.

Hyman and Monaka end their chapter by raising the question of what exactly is intonation, and whether the non-tonal strategies observed in Shekgalagari to mark sentence types can be also considered intonational, given the fact that one of the main functions of intonation is to mark sentence types. This is a relevant question, from a theoretical point of view. If it is true that one of the main functions of intonation is to signal sentence type, it is also true that it is only one of several other functions, such as phrasing utterances into prosodic constituents, signalling focus, or expressing paralinguistic meaning. Hence an important issue for discussion in cross-linguistic research is whether it is appropriate to refer to the segmental marking of sentence types as *intonational*, or to restrict the term intonation to categories conveyed by suprasegmental cues.

The collection of studies presented in this volume will be of interest to a broad range of linguists and language researchers, such as phoneticians, phonologists, morphologists, syntacticians and semanticists interested in the syntax-phonology interface and the import of prosody to pragmatics and semantics. It will also be of interest to speech scientists, and to those with an interest on psycholinguistics and language acquisition and development.

We would like to thank all the contributors for their articles, and all the scholars who had agreed to review the chapters submitted to this book, as well as the helpful advice received from Springer's anonymous referees. Thanks are also due to Marisa Cruz for precious help in the preparation of the final book manuscript. This work has been partially funded by grants from Project FFI2009-07648/FILO and CONSOLIDER-INGENIO 2010 Programme CSD2007-00012 (both awarded by the Ministerio de Ciencia e Innovación, Spain), Project 2009 SGR 701 (awarded by the Generalitat de Catalunya, Spain), Project PTDC/CLE-LIN/108722/2008 (awarded by Fundação para a Ciência e a Tecnologia, Portugal), and by financial aid to the Research Group in Theoretical Linguistics HiTT (given by the Basque Government, Spain, ref. GIC07/144-IT-210-07) and to the Center of Linguistics of the University of Lisbon (given by Fundação para a Ciência e a Tecnologia, Portugal).

References

Behrens, H., and U. Gut. 2005. The relationship between prosodic and syntactic organization in early multiword speech. *Journal of Child Language* 32: 1–34.

Borràs-Comes, Joan, Maria del Mar Vanrell Bosch, and Pilar Prieto. 2010. role of pitch range in establishing intonational contrasts in Catalan. In *Proceedings of Speech Prosody 2010*, Chicago (http://speechprosody2010.illinois.edu/papers/100103.pdf).

Carlson, Katy, Charles Clifton, Jr. and Lyn Frazier. 2001. Prosodic boundaries in adjunct attachment. *Journal of Memory and Language* 45: 58–81.

Chen, Aoju. 2003. Reaction time as an indicator of discrete intonational contrast in English. In *Proceedings of Eurospeech 2003*, Geneva, 97–100.

Chen, Yiya. 2007. The phonetics and phonology of consonant-F0 interaction in Shanghai Chinese. Talk present at the workshop on Where Do Features Come From? *Phonological Primitives in the Brain, the Mouth, and the Ear*. Paris, October 5th, 2007.

Clifton, Charles Jr., Katy Carlson and Lyn Frazier. 2002. Informative prosodic boundaries. *Language and Speech* 45: 87–114.

D'Imperio, Mariapaola. 2000. The Role of Perception in Defining Tonal Targets and their Alignment. Ph.D. Thesis, The Ohio State University.

D'Imperio, Mariapaola, and David House. 1997. Perception of questions and statements in Neapolitan Italian. In *Proceedings of Eurospeech '97*, Rhodes, Greece, vol. 1, 251–254.

Falé, Isabel, and Isabel Hub Faria. 2006. Categorical perception of intonational contrasts in European Portuguese. In *Proceedings of Speech Prosody 2006*, Dresden, 2–5 May 2006 (on CD-ROM).

Féry, Caroline, and Vieri Samek-Lodovici. 2006. Focus projection and prosodic prominence in nested foci. *Language* 82(1): 131–150.

Fodor, Janet Dean. 1998. Learning to parse? *Journal of Psycholinguistic Research* 27: 285–319.

Fodor, Janet Dean. 2002. Psycholinguistics cannot escape prosody. In *Proceedings of Speech Prosody*, 83–88. Aix-en-Provence, France.

Frota, Sónia. 2000. *Prosody and Focus in European Portuguese. Phonological Phrasing and Intonation*. New York: Garland Publishing.

Frota, Sónia. 2002. The Prosody of Focus: a Case-Study with Cross-Linguistic Implications. In *Proceedings of Speech Prosody 2002*, Aix en Provence, 319–322.

Frota, Sónia. 2010a. A focus intonational morpheme in European Portuguese: Production and Perception. In Gorka Elordieta and Pilar Prieto (eds.) *Prosody and Meaning*. Berlin: Mouton de Gruyter, submitted.

Frota, Sónia. 2010b. Prosodic structure in early child speech: Evidence from intonation, tempo and coda production. Talk present at the Workshop on Prosodic Development. Universitat Pompeu Fabra, Barcelona (http://www.fl.ul.pt/LaboratorioFonetica/texts/WPD_Frota2010.pdf).

Hallé, P., B. Boysson-Bardies, and M. Vihman. 1991. Beginnings of prosodic organization: intonation and duration patterns of disyllables produced by Japanese and French infants. *Language and Speech* 34(4): 299–318.

Harris, Martin. 1988. French. In Martin Harris and Nigel Vicent (eds.) *The Romance Languages*, 209–245. London: Routledge.

Hualde, José Ignacio. 2006/2007. Stress removal and stress addition in Spanish. *Journal of Portuguese Linguistics* 5–2/6–1: 59–89.

Jessen, Michael and Justus C. Roux. 2002. Voice quality differences associated with stops and clicks in Xhosa. *Journal of Phonetics* 30: 1–52.

Jun, Sun-Ah. 2003. Prosodic phrasing and attachment preferences. *Journal of Psycholinguistic Research* 32: 219–249.

Jun, Sun-Ah. 2005. Korean Intonational Phonology and Prosodic Transcription, Prosodic Typology. In Sun-Ah Jun (ed.) *Prosodic Typology. The Phonology of Intonation and Phrasing*, 201–229. Oxford: Oxford University Press.

Keating, Patricia A. 1988. The phonology-phonetics interface. In Frederik J. Newmeyer (ed.) *Linguistics: The Cambridge Survey*, vol. I. *Linguistic Theory: Foundations*, 281–302. Cambridge: Cambridge University Press.

Kenstowicz, Michael. 1994. *Phonology in Generative Grammar*. Cambridge, Mass.: Blackwell.

Kingston, J. and R. L. Diehl. 1994. Phonetic knowledge. *Language* 70: 419–454.

Peperkamp, Sharon. 1997. *Prosodic Words*. HIL Dissertations 34. The Hague: Holland Academic Graphics.

Prieto, Pilar, Eva Estebas-Vilaplana, and Maria del Mar Vanrell Bosch. 2010. The relevance of prosodic structure in tonal articulation. Edge effects at the prosodic word level in Catalan and Spanish. *Journal of Phonetics*, 38/4.

Rialland, Annie, and Stéphane Robert. 2001. The intonational system of Wolof. *Linguistics* 39(5): 893–939.

Schneider, Katrin, Grzegorz Dogil, and Bernd Mobius. 2009. German boundary tones show categorical perception and a perceptual magnet effect when presented in different contexts. In *Proceedings of Interspeech*, Brighton, 2519–2522.

Vigário, Marina. 2003. *The Prosodic Word in European Portuguese*. Berlin/New York: Mouton de Gruyter.

Watson, Duane and Edward Gibson. 2005. Intonational phrasing and constituency in language production and comprehension. *Studia Linguistica* 59: 279–300.

Wheeler, Max. W. 2005. *The Phonology of Catalan*. Oxford: Oxford University Press.

Phonological Trochaic Grouping in Language Planning and Language Change

Aditi Lahiri and Linda Wheeldon

1 Grouping of Morphosyntactic and Phonological Constituents

Grouping (constituency) is our only concern here. How do the parts and wholes of morphosyntactic constructions relate to the parts and wholes of phonological constructions? Morphosyntactic constituents are largely, though not always, meaningful (morphemes, morphosyntactic words, morphosyntactic phrases, sentences etc.) and phonological constituents are largely meaningless (features, segments, syllables, feet, phonological words, phonological phrases, etc.). An analogous question concerning the relation between semantic and morphosyntactic grouping can be raised. The grouping in constructing semantic representations tends to determine/to be mirrored by morphosyntactic grouping (compositionality), but does not have to be (cf. Lahiri and Plank 2007, in press; Plank and Lahiri 2009). Example (1) points to possible obstacles for transparent mapping. The morphological bracketing in (1a) is imposed by English grammar, which permits the negative prefix un- to combine only with adjectives to form adjectives (and rarely verbs). However, on purely semantic grounds a grouping of a negative with a nominal would be equally plausible as in (1b). The relevant grouping for lexical phonological domains as stress and *Trisyllabic Shortening* are again not isomorphic to the others (1c), where -*ity* must be suffixed before the prefixation of *un*-. Finally, syllabic and morphological bracketing rarely coincides as in (1d).

(1) Constituent grouping

(a)	[[un-[[de-[cipherN]V]-abilA]A] -ity N]	morphological
(b)	[[un-] [[[de-[cipherN]V]-abilA]-ityN] N]	semantic
(c)	[[[un-] [[de-[cipherN]V]-abilA]A]-ityN]	lexical phonological domains
(d)	(un).((de.CI.phe.ra).(BI.li.ty))	postlexical phonological grouping

A. Lahiri (✉)
Faculty of Linguistics, Philology and Phonetics, University of Oxford, Oxford, UK
e-mail: aditi.lahiri@ling-phil.ox.ac.uk

S. Frota et al. (eds.), *Prosodic Categories: Production, Perception and Comprehension*, Studies in Natural Language and Linguistic Theory 82,
DOI 10.1007/978-94-007-0137-3_2, © Springer Science+Business Media B.V. 2011

Note that it is not so much the divisions into units — morphemes vs. syllables — which diverge, but their groupings which diverge indiscriminately.

Our focus in this paper is on the grouping of units on the level of utterances. Is there a natural, preferred grouping in certain languages, of lexical words into larger constituents based on rhythmic principles? If yes, what evidence do we have for such grouping? Scholars such as Henry Sweet in the mid 1800s did recommend such grouping when explaining constituency in natural speech in the course of second language learning. He provided a trochaic tone-group of English texts for the benefit of German language learners as shown in (2) (from Lahiri and Plank 2007, in press).

(2) Henry Sweet's (1885) trochaic grouping

(a) in conventional English orthography
People **used to** $_{\text{(FOCUS)}}$ think the earth was a flat cake, with the sea all round it; but we know now that it's really round.

(b) syntactic bracketing
[[people] [[[used] [to think]] [[[[the] [earth]] [[was] [[a kind] [[of] [[flat] [cake]]]]]], [[with] [[[the] [sea]] [[all] [[round] [it]]]]]]]]

(c) Trochaic grouping
(people) (**used to** think the) (earth) (was a flat cake), (with the) (sea all) (round it) (but we) (know now) (that it's) (really) (round).

(d) IPA variant of Sweet's Broad Romic
-pijpl juwsttəþiŋkði əəð wəzəkaindəv flæt keik̀, -wiðə sijɔl raundit bətwij nou nau ðætits riəli ˑraunđ

The grouping in (2d) shows that the syntactic structure is maintained only when it follows the trochaic grouping. Preposition and determiners are grouped together (e.g. *with the*) ignoring the syntactic phrasing. Likewise, *used to think the* is a perfectly unaffected grouping in a normal speech production, but cannot be interpreted as a meaningful constituent. Sweet provides comparable phonological grouping of Dutch sequences as well.

(3) Dutch syntactic vs. trochaic grouping following Sweet (1885)

(i)	(Geef mij en licht) give me a light	(Geef mij en) (licht)
(ii)	Heb je goed geslapen? have you good sleep-PART	(Heb je) (goed ge) (slapen)? [xutxe]
(iii)	Ik kan mijn boek niet finden I can my book not find-INF	(Ik kan mijn) (boek niet) (finden)

Thus, Sweet's instruction to foreigners suggests that at least in Germanic, the phrasal rhythm as he perceived it a century ago was trochaic, where a strong stress prefers to attract upcoming unstressed elements. Evidence for leftwards attachment which lead to cliticization and eventually grammaticalised suffixes is available from most languages (e.g. Plank 2005 for Latin). We will briefly discuss the effect of such cliticization from Swedish, Norwegian, and Bengali, where trochaic grouping has led to the creation of single phonological words.

Other than leftwards attachment of unstressed items and their formations into prosodic words, there is another type of trochaic grouping which is pertinent — the case of compounds. If we consider ordinary two-word lexical compounds, like *blackbird*, then a compelling assumption is that the compound is a syntactic word consisting of two prosodic words but with one main stress. Thus, *blackbird* is a noun, with main stress on the first foot. How does one then prosodically label a compound? Do we allow recursive prosodic word formation and call it a single prosodic word (Lahiri, Jongman and Sereno 1990)? In the prosodic hierarchy literature we do not have a category that fits naturally. Under Selkirk's (1986) assumption, a compound could be a *minor phrase*. Booij (1995) proposed a recursive structure where compounds could be assumed to be a prosodic word where one prosodic word is Chomsky-adjoined to a preceding prosodic word. Under Nespor and Vogel's (1986) hierarchy, we have the choice of a *Clitic Group* but this would be a misnomer. More recently, Vogel and Wheeldon (in prep) and Vogel (2009, submitted) have proposed a *Composite Group*, while Vigário (2009) introduced the *Prosodic Word Group*, both of which essentially replaces the *Clitic Group* as an alternative to recursive phonological word formation. Regardless of the recursive/non-recursive debate, the trochaic grouping remains the same.

The hypothesis we entertain is that normal trochaic grouping and compounding lead to prosodic constituents which play a vital role in normal language production planning and lead to change in morphophonological systems. By language production, we mean not only the acoustic or articulatory outputs but the processing involved in planning to produce speech (cf. Levelt 1989). What are the units used by speakers to plan their articulation? Are they lexical words or are they prosodically grouped structures not necessarily isomorphic to syntactic structure? We begin with a discussion of examples of trochaic grouping which led to cliticization and suffixation in Scandinavian (Section 2), Germanic and Bengali verb morphology (Section 3) and English cliticization (Section 4). In Section 5, we show that different types of prosodic structures based on trochaic grouping can form different domains for different rules. The examples will be from Dutch. We then turn to psycholinguistic evidence (Section 6) arguing that language planning does not involve simple lexical words but rather prosodic structures related to trochaic grouping.

2 Leftwards Grouping with DEFINITE ARTICLE: Evidence from Swedish and Norwegian

Certain Swedish and Norwegian dialects maintain contrastive tonal accents *villa*$_1$ 'villa' vs. *flicka*$_2$ 'girl'. A striking contrast that has emerged in these languages involves the definite article which attaches leftwards to nouns. Phonologically they are quite distinct from, for instance, plural suffixes. This cliticization, to a large extent, consolidated the accent contrast in Scandinavian: *lagen*$_1$ vs. *lagar*$_2$ 'the law/laws'. The synchronic syntactic structures often show double determination, which is not unusual in other languages (Plank 2003). Other than Danish, double determination is the default rather than the exception (Börjars 1994).

(4) Scandinavian determiners

(a) Swedish: must co-occur with one exception (*denna* 'this' cannot co-occur in most dialects)
den gamla mus-**en**/*mus
den mus-**en**/*mus
den där mus-**en**/*mus
denna ?*mus-**en**/mus

(b) Norwegian: must co-occur (even *denne* must co-occur)
same as Swedish except:
denne bil-**en**

(c) Danish: def occurs in complementary distribution with determiners:
mand-**en**
den mand/*mand-**en**
denne mand/*mand-**en**

The argument from syntacticians is that the determiner is an affix rather than a clitic since it is placed in affix-typical position. Here the DEF attaches to the first word of the phrase only when there is no premodification; to the final word only when there is no postmodification (Lahiri, Wetterlin and Jönsson-Steiner 2005a). However, this controversy is not relevant. What is crucial is the grouping which led to the definite marker being attached to the noun. The article derives from the demonstratives in 'articular' (=joining) function with an attributive adjective following after a noun which was the normal adjective position of old. This is the same in most Germanic languages. Thus the constructions that would have led to the grouping would be of the following type:

(5) Steps in double definiteness due to leftwards attachment in Scandinavian

(i) warrior, this/the valiant [one]
(ii) warrior=the valiant
(iii) valiant warrior=the
(iv) the valiant warrior=the

The effect of the trochaic grouping is clearly seen through the proposed stages in (5). From the construction in (5i), the article was prosodically rephrased and reanalysed with the preceding stressed noun as in (5ii). And this is where the article remained when nouns were on their own, once definiteness marking had become obligatory (5iii). When the regular adjective position came to be pre-nominal, adjectives took the definite marker with them.

In these synchronic systems, one can still see the consequences of the cliticization in the tonal outputs. When one attaches the indefinite plural suffix to a monosyllabic noun, it gets assigned Accent 2 as a normal trochaic word would. If, instead, the definite ending is attached, also forming a trochee, the noun remains Accent 1. Indeed the definite ending has no effect on the tonal properties of the noun and it thus behaves phonologically like a clitic (Riad 1998; Lahiri, Wetterlin and Jönsson-Steiner 2005b; Kristoffersen 2000, and references therein).

(6) Definite clitics and plural suffixes in Scandinavian

	Swedish	*Norwegian*
sg	stol$_1$ månad$_2$ te̽rmos$_1$	stol$_1$ måned$_2$ te̽rmos$_1$
sg.def.	stol=en$_1$ månad=en$_2$ te̽rmos=en$_1$	stol=en$_1$ måned=en$_2$ te̽rmos=en$_1$
pl	stol-ar$_2$ månad-er$_2$ te̽rmos-ar$_1$	stol-er$_2$ måned-er$_2$ te̽rmos-er$_1$
pl.def.	stol-ar=a$_2$ månad-er=na$_2$ te̽rmos-ar=na$_1$	stol-er=ne$_2$ månad-e(r)=ne$_2$ te̽rmos-e(r)=ne$_1$
Gloss	chair/chair=DEF/chair-PL month/month=DEF/month-PL thermos/thermos=DEF/thermos-PL	chair/chair=DEF/chair-PL month/month=DEF/month-PL thermos/thermos=DEF/thermos-PL

The standard assumption is that the definite clitic is attached to the prosodic word after accent assignment, while the plural suffix is attached before. In the nouns above, the word *te̽rmos* is marked with an asterisk to indicate that it is lexically specified for Accent 1 (Lahiri et al. 2005a). We turn to this below.

(7) Attachment of plural suffix and definite clitic

[/stem/-(PL)$_{\text{accent}}$]ω =DEF

Swedish plural & definite	/stol/-ar	/månad/-er	/te̽rmos/-ar
accent assignment	stolar$_2$ [stolar$_2$]ω=DEF *stolarna*$_2$	månader$_2$ [månader$_2$]ω=DEF *månaderna*$_2$	te̽rmosar$_1$ [te̽rmosar$_1$]ω=DEF *termosarna*$_1$

Norwegian

singular definite	/stol/	/månad/	/teͯrmos/
accent assignment	$stol_1$	$månad_2$	$teͯrmos_1$
	$[stol_1]_\omega{=}DEF_1$	$[månad_2]_\omega{=}DEF_2$	$[teͯrmos]_{1\omega}{=}DEF_1$
	stolen$_1$	*månaden*$_2$	*termosen*$_1$

According to Lahiri et al. (2005b), words like *termos* are represented with their accent in the lexicon, indicated here with the accent mark (*). Specified lexical accent is always interpreted as Accent 1 and overrides the default accent assignment rule. Consequently, irrespective of whether the plural suffix or the clitic is added, words like *termos* always have Accent 1. The default rule states that trochaic words, if unspecified for lexical accent, bear Accent 2. All specified words have Accent 1, including all words that do not consist of a trochee, e.g. all monosyllabic words. A word like *stol*, which is not specified for underlying accent, is assigned Accent 1 since it is monosyllabic. After plural suffixation, that leads to a trochaic structure it is assigned Accent 2 e.g *stolar*. However, it is already Accent 1 in the singular form when the definite clitic is attached. In other words, the clitic has no effect on accentuation.

(8) Accent assignment

(i) Lexically specified (indicated by $^\times$) are always assigned Accent 1
(ii) If no specification and the word has a trochee (*kirke*) then it is assigned Accent 2, else Accent 1 (which includes monosyllabic words)

Since we are also interested in the prosodic structure of compounds, we can note that the clitic attached to a compound again has no affect on the accentuation. Here we only refer to Norwegian, specifically Standard East Norwegian, since Swedish compounds all bear Accent 2. Compound accent assignment in Norwegian is sensitive to the lexical accent of the first prosodic word of the compound as we see in the examples in (8) (Wetterlin 2008).

(9) Compound accent in Standard East Norwegian

$[\overset{\times}{\omega}\ \ \omega\,]$ > Accent 1 (first word is lexically specified)

else, $[\,\omega\ \omega]$ > Accent 2 (as in any default trochaic accent)

Lexical repr. ω_1	/kirke/		/aͯksje/	
Lexical repr. ω_2	/tårn/	/oͯrgel/	/bank/	/marked/
Compound	'kirkeˌtårn$_2$	'kirkeˌorgel$_2$	'aksjeˌbank$_1$	'aksjeˌmarked$_1$
Gloss of compound	*church tower*	*church organ*	*stock bank*	*stock market*
DEFINITE	'kirkeˌtårn=et$_2$	'kirkeˌorgel=en$_2$	'aksjeˌbank=en$_1$	'aksjeˌmarked=et$_1$

The word *aksje* is lexically specified for Accent 1 and this determines the accent of the compound. When the accent of the word is unspecified as in *kirke* (it would take default Accent 2 in isolation because it is a trochee), the compound as whole bears Accent 2. What is crucial here is that accent assignment of the compound must come after compounding, as we can see from the following examples where the first prosodic word is monosyllabic. Although monosyllabic words in isolation are always Accent 1 (see 8), they influence compounds in different ways. We follow Lahiri et al. (2005a) and Wetterlin (2008) in assuming that the difference lies in the lexical accent specification; some monosyllables are specified for Accent 1, and some are not.

(10) Monosyllabic first word with Accent 1 and 2

Lexical repr. ω_1	/land/		/skŏ̽/	
Lexical repr. ω_2	/vei/	/tunge/	/krem/	/fa'brikk/
Compound	'landvei$_2$	'landˌtunge$_2$	'skoˌkrem$_1$	'skofaˌbrikk$_1$
Gloss of compound	*country path*	*peninsula*	*shoe cream*	*shoe factory*
DEFINITE	landvei=en$_2$	'landˌtunge=en$_2$	'skoˌkrem=en$_1$	'skofaˌbrikk=en$_1$

The monosyllabic *land* is not specified with its accent in the underlying representation. It gets Accent 1 as default when uttered in isolation since it is monosyllabic. However, since it is unspecified for accent, it has no effect on the compound accent which is the default (trochaic) Accent 2. Thus, *land* must get its accent <u>after</u> compounding, and the cliticized *land = en* is formed <u>after</u> accent assignment. The noun *sko*,* on the other hand, is specified for Accent 1, and hence its accent has an effect on the compound accent which is assigned Accent 1. Note that accent assignment is not influenced by the accent of the second member. The definite article cliticizes leftwards to attach to the compound. Again these are not suffixal since they remain outside the tonal domain. Consequently, the cliticization of the definite article is exactly the same for compounds and single prosodic words. Our assumptions of compounding and accent assignment follow Wetterlin (2008). Examples in (11) exemplify the interaction of accent assignment and compounding.

(11) Accent assignment and compounding

	/land/	/skŏ̽/	/land/	/skŏ̽/
compounding	(land)(vei)	(skŏ̽)(krem)		
lexical & compound accent assignment	(landvei)$_2$	(skokrem)$_1$		sko$_1$
default accent assignment			land$_1$	
cliticization	((landvei)$_2$=en)$_2$	((skokrem)$_1$=en)$_1$	(land)$_1$=en)$_1$	(sko)$_1$=en)$_1$

What we cannot convincingly determine from these compounds is whether they are definitively formed on two prosodic words or whether they are still stems which are combined to make a compound. This is unlikely since one usually assumes that compounds are made up of two words, i.e. $((\omega)(\omega))\omega_{\text{compound}}$. The clitics are attached to the entire compound which is a prosodic word on its own and therefore we will have a recursive formation, viz. $(((\omega)(\omega))_{\omega\text{-(compound)}}=\text{CLITIC})_{\omega}$. If we assume no recursivity, and consider compounds to be *Composite Group*, what would be the constituent after cliticization: e.g. $(((\omega)(\omega))_{\text{CG(compound)}}=\text{CLITIC})_{X}$? Clearer evidence for an independent compound domain comes from Dutch (Section 5). Before that we touch briefly on Bengali and Germanic auxiliaries, which cliticized to verbs, again providing evidence for trochaic grouping.

3 Germanic and Bengali Auxiliary Cliticization in Verbs

The verb 'do' provided tense marking in weak verbs in all Germanic languages. In Bengali the auxiliary 'be' has provided a progressive suffix. Both have been consequences of the main verb roots attracting the less strong auxiliary in its prosodic domain.

3.1 Germanic weak verbs were made from nouns, adjectives or other verbs with the addition of suffixes, most commonly the /j/ causative suffix which caused gemination and umlaut in certain conditions. Ablaut, or changing the root vowel under specific morphological conditions was the dominant way of marking the past; Old English *helpan, healp, hulpon, holpen* 'help INF, 1/3SG PAST INDIC, PLURAL PAST INDIC, PAST PARTICIPLE.The modern Germanic languages all maintain a present-past distinction in these verbs; cf. English *come, came;* German *komme, kam*; Dutch *kom, kam* etc. Ablaut was, however, not possible to indicate the past tense of derived verbs, since the root vowel in most instances would have been umlauted and therefore a front vowel. The vowel alternation pattern of ablaut verbs was not available. Consequently, the past was constructed as a compound verb by adding the past of 'to do' (Lahiri 2000).

(12) Morphological decomposition of derived weak verbs in older Germanic

Present tense [ROOT+ /j/ $_{\text{CAUSATIVE SUFFIX}}$] + PERSON-NUMBER inflection
Past tense - compound formation, where X=infinitive or verbal noun
[ROOT + /j/ $_{\text{CAUSATIVE SUFFIX}}$ + X]ω + [*do*$_{\text{AUX}}$-PAST-PERSON-NUMBER inflection]

> [[ROOT]ω =*do*$_{\text{AUX}}$-PAST-PERSON-NUMBER inflection]ω
> [[ROOT] -*d* - PAST-PERSON-NUMBER inflection]ω

For instance, a word like *fall* would be made causative with the suffix /j/ which would trigger umlaut and generate *fell*. Present tense suffixes could be added to it like *he fells the tree*. But for the past tense, one would need to make a construction as in *fell did* which later became *felled*.The strong past patterns as in *ring-rang* could not be used as a template.

The modern Germanic languages all have the coronal stop /t/ or /d/ as the past marker. Whether they are voiced or voiceless depends on the normal historical development and assimilation; English {d,t,ed}, German {t, et}, Dutch {d, t, et} etc. For verbs ending with a coronal stop, either a schwa is inserted between the root and the coronal stop or the stop is deleted.

(13) Past tense in modern Germanic weak verbs

[ROOT+ coronal stop] + person-number inflection

	INF	PAST-3p	INF	PAST-3p
English	pat	patt-ed [ed]	beg	begged[d]
German	red-en	red-ete [ete]	schraub-en	schraub-te[t]
Dutch	rijd-en	reed	krabb-en	krab-de[d]

	INF	PAST-3p
English	kiss	kiss-ed [t]
German	hüpf-en	hüpf-te [t]
Dutch	knijp-en	knijp-te[t]

Thus, the past tense of *do*/*tun*/*doen* cliticized to the root. The coronal stop of *do* became a morpheme indicating past. What is consistent is that the coronal stop has been retained, while the inflectional suffixes conform to the morphological system of the language. The suffixed forms are prosodic words in their entirety. The verb ‘do’, has continued to exist as an independent verb. Again, we see a trochaic grouping causing leftwards attachment of the auxiliary.

3.2 The auxiliary /ach-/ ‘to be’ in Bengali is suppletive; only the present and the simple past tense forms exist. Through the last 1000 years it has been used to supplement the verbal aspectual system. One example is that of the present progressive which is a suffix derived from the original /ach-/. The suffix consists of the palatoalveolar consonant, but has become underlyingly a geminate due to regular sound changes Lahiri (2000), Lahiri and Fitzpatrick-Cole 1999.

(14) Bengali progressive forms

	ROOT-1PERS.PRESENT	ROOT-PROGRESSIVE-1PERS
lie down	shu-i	shu-cch-i
play	khel-i	khel-ch-i

The progressive developed from a full verb form and gradually was attached leftwards to the root and lost its vowel. Later, in the context of vowel final roots, the palatoalveolar consonant geminated and is now the underlying form of the suffix.

(15) Development of the progressive

$[\text{ROOT - PROG}]_{\omega}\ [ach_{\text{AUX}}\text{-}\ e_{\text{ PRESENT-3PERSON}}]_{\omega}$

$[\text{ROOT - PROG}]_{\omega}\text{=}\ [ch_{\text{AUX}}\text{-}\ e_{\text{ PRESENT-3PERSON}}]_{\omega}$

$[\text{ROOT - }cch_{\text{ AUX}}\text{-}\ e_{\text{ PRESENT-3PERSON}}]_{\omega}$

	lie down	play
Old Bengali	shu-ite ach-e	khel-ite ach-e
Middle Bengali	shu-i ach-e	khel-i ache-e
Early modern	shu-i=ch-e	khel-i=ch-e
Modern	shu-cch-e	khel-ch-e

An interesting factor is the underlying geminate in the new progressive morpheme /-cch/. This was an innovation, although medial geminates were frequent in the language, and no inflectional morpheme had as yet an underlying geminate. Again, for our purposes, we find the leftwards attachment of an auxiliary to make a trochaic grouping.

4 Cliticizations of Unstressed Words in English

The classic example non-isomorphism between syntactic phrasing and phonological phrasing has always been embedded structures. The syntactic structure has four levels of embedding, while the phonological grouping is flat as we can see in (16).

(16) Non-isomorphism between syntactic and phonological phrasing

(i) [[[[The cat] [[that ate the mouse] [[that ate the cheese] was sick.]
(ii) (The cat that) (ate the mouse that) (ate the cheese) (was sick)

Selkirk (1995, 1996) provides a full list of possibilities of phonological groupings with function words. She assumes that above a (morphological) word, there is the assumption of a necessarily close match between syntactic and phonological mapping, in that phonological grouping is in essence determined by (morpho)syntactic grouping. The possible prosodizations of English [Fnc Lex] with [XP] and phonological phrases (PPh) are as follows, with brackets/parentheses coinciding owing to a general constraint on Edge Alignment such that XP/PPh brackets coincide.

(17) Selkirk's function words prosodization

S-Structure	[Fnc Lex] XP	
P-Structure	(i) $((\text{fnc})_{\text{PWd}}\ (\text{lex})_{\text{Pwd}})_{\text{PPh}}$	Prosodic Word [function word is not weak)
	(ii) $(\text{fnc}\ (\text{lex})_{\text{Pwd}})_{\text{PPh}}$	prosodic clitic = free clitic
	(iii) $(\text{fnc lex})_{\text{Pwd}})_{\text{PPh}}$	internal clitic
	(iv) $((\text{fnc}\ (\text{lex})_{\text{Pwd}})_{\text{PWd}})_{\text{PPh}}$	affixal clitic

According to Selkirk, weak forms of function words in English appear when non-focused, when not phrase final, and when phrase final, but not as an object of a verb (e.g *Where have you got* ***to****?*). Examples of weak function words (underlined) and their subsequent phrasing following Selkirk are given in (18).

(18) [Diane] [can paint] [her portrait] [of Timothy] [at home]
[But ǂshe found] [that ǂthe weather] [was ǂtoo hot] [for painting]

In English, function words with a weak form in this kind of examples are proclitic (underlined), of the subtype ‘free (pro-)clitics’: *I must fly to Toronto* $((\text{to})_{\text{clitic}}(\text{Toronto})_{\omega})_{\Phi}$. Selkirk’s evidence comes from postlexical rules like aspiration of initial voiceless stops (in stressed as well as unstressed syllables) which is P-Wd-initial; hence *to* T^h*oronto*, *t^h*o Toronto*, *t^h*o* T^h*oronto*. Nevertheless, one can also obtain these facts by trochaic grouping as in (19) (Plank and Lahiri 2009, in press).

(19) Trochaic grouping
[fly to]$_{\omega}$ [Toronto]$_{\omega}$

As we have seen from Sweet’s examples in (2), English also has enclitics, which are weak forms of function words in constructions with lexical words preceding them (e.g., [feed ’em], [see ya]). These are of the subtype ‘affixal (en-)clitics’ which include object pronouns. Additionally, in the Selkirk 1995 approach, there are also *enclitics* in English whose hosts are preceding Lex words:

(20) [Nina] [’s left]; [Mary] [’s coming]; [I] [’ll leave] too; [I] [’d like] [to stay]

Following Lahiri and Plank (2007) and Plank and Lahiri (2009), the hypothesis entertained here is that the default phrasing is left attachment, i.e. encliticization.

(21) Encliticization following trochaic grouping

[Nina has] [left]
[John] [walked to] [school]
[I’d] [like to] [stay]

The complementizer *to*, indeed, is a notorious ‘misfits’ *liketa*, *hafta*, *wanna*, *gonna* etc. which have been discussed a lot in the syntax literature (cf. Zwicky and Pullum 1983) and leftwards attachment is the only explanation. Again, English also has evidence for encliticizations.

5 Encliticization in Dutch

Dutch is no different from other Germanic languages in that function words do not count as phonological words unless focused. Earlier work on Dutch has established that the definite article can easily cliticize leftwards to attach to the preceding verb giving us the familiar grouping in (22).

(22) What is the definite article phonologically grouped with — noun or verb?

syntactic grouping:	VERB [DEF N]
phonological grouping:	(VERB DEF) N

Accepting this phonological grouping, Gussenhoven (1983) proposed a P-word formation (Left; X^0) giving us the grouping in (23).

(23) P-word formation in Dutch

Ik zoek de krant	$_{\omega}$(ik $_{\omega}$(zoek de $_{\omega}$(krant)))
	'I am looking for the newspaper'
syntactic phrasing	(ik) ((zoek) (de krant))

In Dutch, we can obtain evidence from voicing assimilation for the different domains. Here we can directly compare across word assimilations between compounds, cliticized words and a sequence of X^0 categories. Compare the following examples from Lahiri, Jongman and Sereno (1990) (based on Berendsen (1986) and Zonneveld (1983)) indicating the differences between compounds, cliticized prosodic words and across prosodic words within and across a phonological phrase. Here we see clear evidence for a difference between compounds and two separate prosodic words.

(24) Voicing assimilation in different domains

a)	compounds $((\omega\ )\ (\omega\ ))_{\omega\ \text{COMPOUND}}$	regressive assimilation obligatory
b)	P-wds across phrases $((\omega\ ))_{\phi}\ (\omega\ ))_{\phi}$	regressive assimilation optional
c)	cliticized word $((\omega\ ) = \text{Fnc}_{\text{CLITIC}}\ )_{\omega}$	regressive or progressive assimilation optional

The optionality and obligatory character of the voicing assimilations are made explicit in the following examples.

(25) Optionality of voicing assimilations across lexical boundaries

<table>
<tr><td rowspan="2">underlying</td><td>a. $((\text{meet})_{\omega}\ (\text{band})_{\omega})_{\omega}$
measuring tape</td><td>b. ik vind
$((\text{Joop})_{\omega}\ (\text{dun})_{\omega}\ )_{\Phi}$
$((\text{Joop})_{\omega}\ )_{\Phi}\ ((\text{dun})_{\omega}\ )_{\Phi}$
I find Joop thin</td><td>c. ik zoek der (haar)
$((\text{zoek})_{\omega}{=}\text{der})_{\omega}$
$(\text{zoekder})_{\omega}$
I look for her</td></tr>
<tr><td>$((\text{zak})_{\omega}\ (\text{doek})_{\omega})_{\omega}$
handkerchief</td><td></td><td></td></tr>
<tr><td>regressive assimilation</td><td>mee[d][b]and

za[g][d]oek</td><td>joo[b][d]un
*joo[b][d]un</td><td>zoe[g][d]er
*zoe[g][d]er</td></tr>
</table>

progressive assimilation	*mee[t][p]and *za[k][t]oek	*joo[p][t]un *joo[p][t]un	*zoe[k][t]er zoe[k][t]er
No change	*mee[t][b]and *za[k][d]oek	*joo[p] [d]un joo[p] [d]un	*zoe[k][d]er *zoe[k][d]er

The clitic *der* cannot be stressed. If it is, then the full pronoun *haar* has to be used. Voicing assimilation is a must in a compound, but not so for the cliticized words. Following cliticization, there are two options: either the clitic joins with the preceding word like a single lexical item and then the constraint for such items comes into play, viz., no voiced clusters word internally (there is probably one exception, *abdomen* (cf. Zonneveld 1983); or *der* cliticizes to the preceding word and undergoes voicing assimilation like a compound. The crucial point is that *der* must share the voicing of the preceding word, and the subsequent cluster must be either voiced or voiceless. For a compound, the sequence has to be voiced if the initial stop of the second word is voiced (Zonneveld 1983). Notice that for (25b), when there are two prosodic words which may or may not be in phonological phrase, voicing assimilation is possible, but not obligatory as it is for the compound. Thus, the cliticized forms, compounds and a sequence of prosodic words are subject to different constraints for voicing assimilation.

Returning to the Gussenhoven's P-word formation in (23), and following Berendsen (1986), we can find additional evidence from voicing assimilation facts. Consider the following possibilities where voicing assimilation allows for two options: [zoekte] or [zoegde].

(26) Voicing assimilations for the cliticized definite article

	Ik zoek de krant		
	Ik (zoe[k] [t]e)$_{\omega}$ (krant)$_{\omega}$	progressive assimilation of def art	1Pwd
	Ik ((zoe[g])$_{\omega}$[d]e))$_{\omega}$ (krant)$_{\omega}$	regressive assimilation of verb	clitic attached to preceding Pwd;
	*Ik (zoe[k])$_{\omega}$ ([d]e krant)$_{\omega}$	no assimilation is not possible	clitic attached to following Pwd
cf.	*ik vind Dik dun*	‘I find Dik thin’	
	Ik vind (Dik)$_{\omega}$ (dun)$_{\omega}$	where [k] [d] as well as [g][d] are possible.	

These postlexical processes are always optional and variable. What we want to note here are the differences between the levels of prosodic grouping. Cliticized words share properties with compounds, but are different in some ways. Similarly, compounds are different from a sequence of two lexical prosodic words which do not form a compound. These differences are more easily accounted for in a system which allows recursive phonological word formation.

Now we turn to online sentence production. Do we have any psycholinguistic evidence from sentence generation — not just measuring the acoustics after production — but the actual planning involved in the production? The process of producing a sentence includes many steps. Depending on circumstances, the speaker may have a certain amount of time to prepare to speak or must begin articulation with little preparation, for example as a reply to an urgent question. What units does the speaker use to plan his/her utterances? Are they syntactic or are they prosodic? If they are prosodic, do they follow our hypothesis of trochaic grouping argued for above? Do speakers treat compounds as two words or one? We discuss briefly some experimental evidence which begins to address these issues.

6 Can We Find Any Psycholinguistic Evidence for Such Structures?

What is the evidence that prosodic structures such as the phonological words described above play a role during language production processes? Prior to the onset of an utterance, a phonological representation must be planned that guides articulation. We have argued that the lexical item is not the optimal unit for the planning of phonological structure or for its subsequent articulation. In this section we summarize the psycholinguistic research that has focused on the relationship between syntactic and prosodic structures during language production. The existing research has used two experimental methodologies to investigate which units are involved in the planning and articulation of speech.

- In the *prepared speech paradigm*, speakers are required to construct an utterance for output and to prepare to say it on a given cue. This is not an unusual situation in language production, as in conversational settings speakers must often wait for an opening in order to produce utterances that they have already planned. In this case, the time it takes for a speaker to initiate articulation should be determined by the structure of the utterance as a whole. In other words, speech onset latency should be a function of the *number of units* in the utterance. The question addressed by this paradigm is what is the nature of the unit that determines speech onset latency?
- In the *on-line speech production paradigm*, speakers must construct and articulate their utterances as quickly as possible. There is a great deal of evidence that, during fluent speech, language is planned incrementally, with minimal units being constructed at a given level of representation prior to the onset of processing at the subsequent level. In other words speakers do not normally wait until they have constructed and entire utterance before they

> begin to speak. Instead they articulate the first unit of an utterance whilst simultaneously planning subsequent units (Kempen and Hoenkamp 1987; Levelt 1989). In this situation the time it takes to initiate speech will be determined by the **size of the first unit** to be produced.

Thus, planning to produce an utterance entails a decision on the part of the speaker whether to spend time preparing the output or whether to begin speaking as soon as possible. Under both conditions, the problem is the preferred minimal unit of output - is it a syntactic phrasal unit or is it a prosodic phrasal unit constructed online as the articulators plan to produce the output? The minimal planning unit affects the two production strategies in different ways.

(27) Prepared versus online planning of speech production

(i) Prepared production is affected by the **number** of units planned
(ii) Online production is influenced by the **size** of the initial planned unit

We discuss each in turn.

6.1 Prepared Speech Production Studies

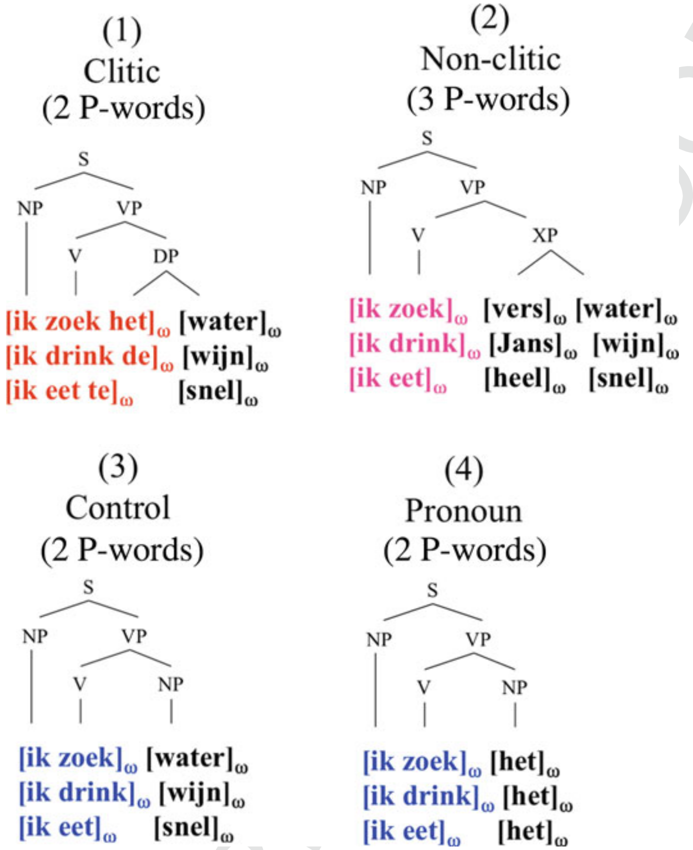
The prepared speech paradigm was first used by Sternberg and colleagues (1978, 1980) to investigate the planning of rapid speech sequences. They asked speakers to prepare to produce lists of random words or digits and to begin producing a sequence at a cued delay. They found that speech onset latencies for the lists increased in a linear fashion as the length of the list increased. In other words, speakers took longer to initiate longer lists. A comparison of different list types helped to determine the nature of the unit that determines list length in this task.

(28)	Monosyllabic words:	bay rum mark
	Disyllabic words:	baby rumble market
	Nouns plus function words:	bay and rum and mark

The critical unit cannot be the number of syllables or lexical items in the list, as the slope of the latency function was the same for monosyllabic and disyllabic word lists, as well as for lists of words including unstressed function words such as ‘and’. The data therefore suggest that all of the list types shown above contain the same number of ‘units’ despite differing in the number of syllables and words. Sternberg et al. concluded that prior to articulation the lists were structured into ‘stress groups’ (e.g., /bay/ /baby/ /bay and/) each of which contained one primary stress and that these units determined list length and therefore speech onset latency in this task.

Wheeldon and Lahiri (1997, 2002) suspected that the units of importance in the Sternberg et al. task might be phonological words. We tested this idea by comparing the delayed production of clitic and non-clitic structures in Dutch.

(29) Test conditions



This figure will be printed in b/w

As can be seen, the clitic (1) and non-clitic (2) sentences are matched for global syntactic complexity and number of lexical items but differ in the number of phonological words they comprise. Therefore, if lexical items are the critical units of production, the latencies to initiate both sentence types should be the same. Alternatively, if phonological words are the critical units then non-clitic sentences should take longer to initiate than clitic sentences.

Two additional sentence types were tested in order to rule out alternative explanations. The control sentences (3) were included to check for effects of the complexity of the initial phonological word. These sentences have the same number of phonological words as the clitic sentences but the initial phonological word is less complex and identical to those of the non-clitic sentences. Any effect of phonological word complexity should be seen in a difference in onset latency between the clitic and control sentences.

Finally, if longer onset latencies are obtained for the non-clitic than the clitic sentences, the effect could be attributed to the fact that the non-clitic sentences contain an additional content word rather than to differences in prosodic structure. The pronoun sentences (4) were included to test this possibility. In these sentences the phrase final determiner attracts stress and therefore becomes a phonological word on its own giving the pronoun and clitic sentences an equal number of phonological words but different numbers of content words. If the number of phonological words rather than number of content words is the critical factor, then onset latencies for these sentence types should not differ.

The order of events on the experimental trials is illustrated below. Speakers were shown the required noun phrase (e.g., het water, *the water*) on a computer screen and then heard a question relating to it (e.g., wat zoek je?, *what do you seek?*). They had a few seconds to prepare their sentences, which they produced following a variable response. Three different cue latencies were used in random order to ensure that speakers could not anticipate when they should start to speak. Although sentence onset latencies were usually shorter following longer cue latencies the pattern of results across cue latencies did not differ. Sentence onset latencies, were measured from the cue to the onset of articulation.

(30) The prepared speech experimental procedure

This figure will be printed in b/w

The results were as we predicted if the critical units in phonological planning are phonological words rather than lexical items.

- Speakers took significantly longer to begin to produce the non-clitic sentences than the clitic sentences.
- The complexity of the initial phonological word did not effect prepared sentence production, as onset latencies for the clitic and control sentences did not differ.
- The number of content words in a sentence had no effect on prepared sentence production, as onset latencies for the clitic and pronoun sentences did not differ.
- Finally, onset latencies were not a function of whole sentence duration, as spoken sentence durations showed a very different pattern of results with all sentence types significantly differing from each other in the direction one would expect.

We concluded that onset latency was a function of the number of phonological words in the utterance.

6.2 On-Line Speech Production Studies

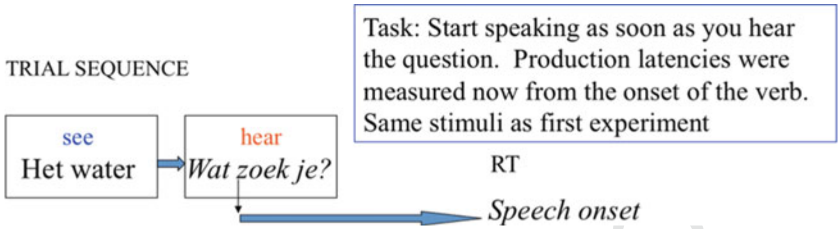
The prepared speech production experiments described in the previous section provide strong evidence that, prior to articulation, stored morpho-lexical representations are restructured into prosodic units. They also provide support for the grouping of unstressed function words into prosodic units. However, these experiments cannot tell us how prosodic structure affects sentence production when the time to prepare is limited. As mentioned above, when planning must occur online, speech is produced incrementally and it is likely that only the minimal production unit is planned prior to the onset of articulation. On-line speech production studies can therefore provide information about the preferred minimal unit of production. If this unit is the phonological word, then the articulators will have to wait for this unit to be planned. In other words, the length of the utterance initial phonological word will determine sentence onset latency (see Levelt 1989; Levelt and Wheeldon 1994).

In addition, while the delayed speech production experiments provide evidence concerning the number of prosodic units constructed, they do not tell us about the direction of attachment during cliticization. We have assumed that the clitic attaches leftwards to the verb, however the right attachment of clitics has also been proposed (Selkirk 1995). Clearly, the direction of attachment will determine the size of the initial phonological word and therefore sentence onset latencies to our clitic sentences. For the sentences given below, left attachment predicts that the clitic sentences should have longer onset latencies that the non-clitic and control sentences. In contrast, right attachment predicts no difference in onset latencies for the three sentence types.

(31)	Clitic, left attachment	[ik zoek het]$_w$ [water]$_w$	2 P-words
	Clitic, right attachment	[ik zoek]$_w$[het water]$_w$	2 P-words
	Non-clitic	[ik zoek]$_w$ [vers]$_w$ [water]$_w$	3 P-words
	Control	[ik zoek]$_w$ [water]$_w$	2 P-words

In order to address these issues, Wheeldon and Lahiri (1997) used the same question-answer technique as we used in our prepared speech experiments but changed the paradigm to elicit on-line sentence production. The timing of events is illustrated below. The critical difference was that speakers were given no time to prepare their responses but had to begin to speak as soon as they could. Sentence onset latencies were measured from the onset of the verb in the question, as the verb is required before sentence construction can begin.

(32) The on-line speech experimental procedure



This figure will be printed in b/w

As predicted, this experiment produced a very different pattern of results to that of the delayed speech production experiments. Onset latencies to the clitic sentences were now significantly slower than onset latencies to the clitic and control sentences, which did not differ. This pattern of results can be explained if we assume that the function words left-attached making the initial phonological word of the clitic sentences longer than those of the non-clitic and control sentences. These data provide support for the proposal that the phonological word is the preferred unit of output during speech production, as speakers clearly prefer to construct such a unit even at the cost of initiation speed. Furthermore, the resultant grouping must be due to encliticization rather than procliticization. That is, the grouping must be *(zoek het) (water)* rather than *(zoek) (het water)*, clearly demonstrating the non-isomorphism between syntactic and phonological structure.

6.3 Compounds vs. Two Prosodic Words

The final question we addressed was whether compounds are treated as one phonological word or two for the purposes of phonological encoding (Wheeldon and Lahiri, 2002). Using a prepared speech task we tested the production of the words and phrases shown below. Each word was produced preceded by the phrase 'het was', *it was*.

(33)	Adj + Noun	[oud] [lid]	Old member
	Compound	[[oog][lid]]	eyelid
	Monomorphemic		
	Initial stress	[orgel]	Organ
	Final stress	[orkest]	Orchestra

This experiment yielded a very clear pattern of results. The production latency for the initial and final stress monomorphemic words did not differ, demonstrating that the location of the stressed syllable was not critical. Critically, the production latencies for the compounds clearly patterned with the morphologically simple words rather than with the adjective–noun phrases whose production latencies were significantly longer that for all other conditions (See also Vogel and Wheeldon, in prep, for a similar pattern of results in Italian). Clearly compounds as compared to phrases function as a single lexical unit with their own lexical meaning.

7 Conclusion

The hypothesis we have been entertaining is that the default grouping of phonological clitics which are usually unstressed function words is trochaic. That is, these clitics attach leftwards to the preceding stressed word. The obvious exception appears to be Romance where trochaic grouping does not hold and where compounds have main prominence on the second element (Peperkamp 1997; Vigário 2003). A line of research worth pursuing is to investigate if there are independent reasons for preferring one grouping over another. This would also have obvious consequences for language acquisition. A further relevant point that has come up in the discussion is recursivity in phonology. In the tradition of Nespor and Vogel (1986), it would be preferable if we could keep phonology distinct from syntax in assuming that there is no recursivity in phonological domains. Consequently a group of two or more phonological words would not constitute another phonological word but rather form a separate phonological level. Vogel (2009) and Vigário (2009) have independently been referring to such a proposal. Let us first consider compounds, which under Nespor and Vogel's analysis, are a problem. They would either have to be fall under a *Phonological Phrase* or would constitute a *Clitic Group,* neither of which are satisfactory. Since a two-word compound would normally consist of two prosodic words, it would be unusual to refer to it as a *Clitic Group* because there are no clitics. Furthermore, a compound has its own lexical properties and could hardly be a phrase. Vogel (2009, in preparation with Wheeldon) has been referring to compounds as a *Composite Group* (CG) which also encompasses phonological cliticisations. It is difficult to see how this would work for recursive three or four word compounds, or could these be designated as phrases?

(34) Two/three word compounds

(a)	high school	$((\text{high})_\omega\ (\text{school})_\omega)_\omega$
(b)	hand ball	$(\text{hand})_\omega(\text{ball})_\omega)_\omega$
(c)	high school handball	$(((\text{high})_\omega\ (\text{school})_\omega)((\text{hand})_\omega\ (\text{ball})_\omega))_\omega$

As we are aware, compounds are extremely problematic and essentially a can of worms. But even simple compounds as those in (34) suggest that these groupings may be more easily explained in a model allowing recursivity in phonological word formation. As the Dutch experimental data show (§6), latencies for compounds were equal to those of monomorphemic words;

i.e. $((\text{oud})_\omega\ (\text{lid})_\omega)_\omega$took the same amount of time to plan as $(\text{orgel})_\omega$. What we did not test was whether $(((\text{high})_\omega\ (\text{school})_\omega)\ ((\text{hand})_\omega\ (\text{ball})_\omega))_\omega$ would also take the same time to plan as a monomorphemic four syllable word $(\text{manufacture})_\omega$. If recursivity was not permitted, (34c) would either have to be a flat structure or a phrase in Vogel's analysis because *highschool* and *handbag* would be CGs and two CGs would make up a phrase.[1] Thus, it is not entirely clear whether the quest of excluding recursivity is a sufficient motive to make the phonological analyses rather complicated.

What is crucial is that experimentally as well as in data from language-change, we have unmistakable evidence that surface morphosyntactic and phonological structure are non-isomorphic. We have provided substantial data of left-attachment in several languages and from various sources — from prescribed pronunciation rules to normal rules of sentence production. Furthermore, language change data also provide additional evidence of encliticization, from North and West Germanic, as well as from Bengali. Finally, psycholinguistic tasks measuring the latency of prepared and online utterances provide additional evidence for leftwards cliticization during sentence generation. Crucially, the data from all of these sources converge on the same trochaic groupings, at least in a subset of languages of the world.

References

Berendsen, E. 1986. *The Phonology of Cliticization*. Dordrecht: Foris Publications.

Booij, G. 1995. *The Phonology of Dutch*. Oxford: Oxford University Press.

Börjars, K. 1994. Swedish Double Determination in a European Typological Perspective, *Nordic Journal of Linguistics* 17: 219–252.

Gussenhoven, C. 1983. Over de fonologie van Nederlandse clitica. (*About the phonology of Dutch clitics*) Spektator 15: 180–200.

Kempen, G., and E. Hoenkamp. 1987. An incremental procedural grammar for sentence formation.*Cognitive Science* 11: 201–258.

Kristoffersen, Gjert. 2000. *The Phonology of Norwegian*. Oxford: Oxford University Press.

Lahiri, A. 2000. Hierarchical restructuring in the creation of verbal morphology in Bengali and Germanic: evidence from phonology. In Lahiri (ed.) *Analogy, Levelling, Markedness,* 71–123, Berlin: Mouton.

Lahiri, A., and J. Fitzpatrick-Cole. 1999. Emphatic clitics in Bengali. In R. Kager and W. Zonneveld (eds.) *Phrasal Phonology*, 119–144, Dordrecht: Foris.

Lahiri, A., and F. Plank. 2007. On phonological grouping in relation to morphosyntactic grouping. Presentation to the *Annual Meeting of the DGfS*, AG 12: Phonological domains: Universals and deviations. Siegen, 28 February – 2 March.

Lahiri, A., A. Jongman, and J. Sereno. 1990. The pronominal clitic [der] in Dutch. *Yearbook of Morphology* 3: 115–127.

Lahiri, A., A. Wetterlin, and E. Jönsson-Steiner. 2005a. Sounds Definite-ly Clitic: Evidence from Scandinavian tone. *Lingue e Linguaggio* IV: 243–262.

[1] One further issue in Vogel's approach is that although recursivity is forbidden, one can skip levels. For example, a prefix need not be a foot, clitic or a word, but may attach directly to the prefixed word.

Lahiri, A., A. Wetterlin, and E. Jönsson-Steiner. 2005b. Lexical Specification of Tone in North Germanic. *Nordic Journal of Linguistics* 28: 61–96.

Lahiri, A. and F. Plank. Phonological phrasing in Germanic: The judgement of history. *Transactions of the Philological Society* in press.

Levelt, W. J. M. 1989.*Speaking: From Intention to Articulation*. Cambridge, MA: MIT Press.

Levelt, W. J. M., and L. R. Wheeldon. 1994. Do speakers have access to a mental syllabary? *Cognition* 50: 239–269.

Nespor, M., and I. Vogel. 1986. *Prosodic Phonology*. Dordrecht, The Netherlands: Foris.

Peperkamp, Sharon. 1997. *Prosodic Words*. HIL Dissertations 34. The Hague: Holland Academic Graphics.

Plank, F. 2003. Double articulation. In F. Plank (ed.) *Noun Phrase Structure in the Languages of Europe*, 337–395. Berlin: Mouton de Gruyter.

Plank, F. 2005. The prosodic contribution of clitics: Focus on Latin. *Lingue e Linguaggio* 4: 281–292.

Plank, F., and A. Lahiri. 2009. When phonological and syntactic phrasing mismatch: Evidence from Germanic. Presentation to *GLAC 15*, Banff, Alberta, Canada, 30 April.

Riad, T. 1998. The origin of Scandinavian tone accents. *Diachronica* XV(1): 63–98.

Selkirk, E. O. 1986. On derived domains in sentence phonology.*Phonology Yearbook* 3, 371–405.

Selkirk, E. O. 1995. The prosodic structure of function words. In J. Beckman, L. Walsh Dickey, and S. Urbanczyk (eds.) *Papers in Optimality Theory, UMASS Occasional Papers in Phonology,* 439–469. Amherst, MA: GLSA.

Selkirk, E.O. 1996. The prosodic structure of function words. In J. Morgan and K. Demuth (eds.) *Signal to Syntax: Bootstrapping from Speech to Grammar in Early Acquisition*, 187–213. Mahwah, NJ: Lawrence Erlbaum.

Sternberg, S., S. Monsell, R. L. Knoll, and C. E. Wright. 1978. The latency and duration of rapid movement sequences: comparisons of speech and typewriting. In G. E. Stelmach (ed.), *Information Processing in Motor Control and Learning*, 117–152. New York: Academic Press.

Sternberg, S., C. E. Wright, R. L. Knoll, and S. Monsell. 1980. Motor programs in rapid speech: Additional evidence. In R. A. Cole (ed.) *The Perception and Production of Fluent Speech*, 507–534. Hillsdale, NJ: Erlbaum.

Sweet, H. 1885. *Elementarbuch des gesprochenen Englisch*. Oxford: Clarendon Press.

Vigário, M. 2003. *The Prosodic Word in European Portuguese*.Berlin: Mouton de Gruyter.

Vigário, M. 2009. The Prosodic Word Group as a domain of prosodic hierarchy. Talk presented at *OCP* 6, Edinburgh.

Vogel, I. 2009. The status of the Clitic Group. In J. Grijzenhout and B. Kabak (eds.) *Phonological Domains: Universals and Deviations*, 15–46. Berlin: Mouton de Gruyter.

Vogel, I., and L. R. Wheeldon. in prep. Units of speech production in Italian.

Wetterlin, A. 2008. *The Lexical Specification of Norwegian Tonal Word Accents*. PhD dissertation. Universität Konstanz.

Wheeldon, L., and A. Lahiri. 1997. Prosodic units in speech production. Journal of Memory and Language 37: 356–381.

Wheeldon, L., and A. Lahiri. 2002. The minimal unit of phonological encoding: prosodic or lexical word. Cognition 85: B31–B41.

Zonneveld, W. 1983. Lexical and phonological properties of Dutch voicing assimilation. In M. van der Broeke, V. van Hoeven and W. Zonneveld (eds.) *Sound Structures,*297–312. Dordrecht: Foris Publications.

Zwicky A., and G. K. Pullum. 1983. Cliticization versus inflection: English n't. *Language* 59: 502–513.

Order Effects in Production and Comprehension of Prosodic Boundaries

Anouschka Foltz, Kristine Maday, and Kiwako Ito

1 Introduction

A vast literature in psycholinguistics deals with the processing of ambiguity in language (see Altmann 1998 for an overview). Studies have focused mainly on lexical and structural ambiguity. One extensively studied structural ambiguity involves a complex noun phrase modified by a relative clause (RC), as in *Someone shot the servant of the actress who was on the balcony* (see Miyamoto 2008 for a list of studies). Here, the RC *who was on the balcony* can either modify the first noun (N1: *servant*) or second noun (N2: *actress*) of the complex NP, giving the respective interpretations that the servant was on the balcony or that the actress was on the balcony. The RC attaches higher in the syntactic tree (*high attachment*) when modifying the first noun than when modifying the second noun (*low attachment*). We will call this construction the RC attachment ambiguity.

This construction has generated interest for two reasons. First, the preferred interpretation varies across languages (Cuetos and Mitchell 1988), posing a problem for universal parsing principles such as Late Closure (Frazier 1978; Kimball 1973). Late Closure proposes that lexical items – if possible – are universally attached into the clause or phrase currently being processed. Here, the second noun is being processed as the parser encounters the RC and thus the RC should be preferably attached to it. Second, the length of the RC affects interpretation preferences, with more high attachment interpretations for long RCs (e.g. Fernández and Bradley 1999 for English). Prosody is claimed to play a crucial role in both of these phenomena (e.g. Fodor 1998, 2002a, 2002b). The cross-linguistic differences in preferred interpretation of ambiguous sentences may emerge as readers adopt language-specific prosodic phrasing during parsing. Also, the length of the RC may

A. Foltz (✉)
Department of Linguistics, Ohio State University, Columbus, Ohio, USA
e-mail: anouschka@ling.ohio-state.edu

S. Frota et al. (eds.), *Prosodic Categories: Production, Perception and Comprehension*, Studies in Natural Language and Linguistic Theory 82,
DOI 10.1007/978-94-007-0137-3_3, © Springer Science+Business Media B.V. 2011

correlate with the likelihood of a prosodic juncture occurring before the RC, which in turn affects the likelihood of attachment choices (cf. Fodor 1998, 2002a, 2002b; Carlson et al. 2001; Clifton, et al. 2002). Since many previous studies on attachment preferences are based on silent reading, where the implicitly produced prosodic patterns are not measurable, the relationship between such *implicit* prosody and the interpretation of sentences has been mostly speculative.

This study focuses on the issue of constituent length and *overt* prosodic phrasing in English – which shows an overall preference for low attachment (Carreiras and Clifton 1993, 1999; Clifton and Frazier 1996; Cuetos and Mitchell 1988; Fernández 2000; Fodor 1998; Frazier 1978). In particular, the study examines (a) whether constituent length affects prosodic phrasing when reading aloud, (b) whether readers' comprehension judgments are consistent with the prosodic phrasings they produce, and (c) whether the order of the two tasks (reading aloud and providing comprehension judgments) modulates these effects.

1.1 Production of Ambiguous Sentences

Previous studies suggest that speakers consistently produce prosodic cues to convey the intended meaning of structurally ambiguous sentences in both conscious disambiguations and task-oriented natural speech. Clifton et al. (2002) have termed this the *rational speaker hypothesis*.

In an early study, Lehiste (1973) instructed speakers to produce sentences containing various structural ambiguities in a way that disambiguated them. Phonetic analyses revealed that speakers used prosody – in particular, lengthening – to disambiguate sentences with more than one possible surface bracketing (e.g. *[Steve or Sam] and Bob will come* vs. *Steve or [Sam and Bob] will come* or *[The old men] and women stayed at home* vs. *The old [men and women] stayed at home*). Listeners also successfully identified the intended interpretations.

Schafer et al. (2000) used a cooperative game task with a predetermined set of utterances to elicit naturalistic speech. Speakers produced temporarily ambiguous sentences such as *When that moves the square will encounter a cookie* (Early Closure construction, where *moves* is intransitive) vs. *When that moves the square it should land in a good spot* (Late Closure construction, where *moves* is transitive). Speakers consistently disambiguated these temporarily ambiguous utterances, even when the context fully disambiguated them: Over 80% of utterances with Early Closure syntax (intransitive verb) had the stronger prosodic boundary after *moves*, and over 70% of utterances with Late Closure syntax (transitive verb) had the stronger boundary after *square*. A following comprehension experiment revealed that listeners successfully determined the intended syntactic structure.

Bradley, Fernández and Taylor (2003) tested how constituent length affects word duration in productions of the RC attachment ambiguity. Speakers produced the target utterances by combining two sentences (1a and b) into a more complex one (1c). Results showed longer N2 (*prince*) durations for sentences with long RCs compared to those with short RCs. The longer word durations suggested that speakers more often placed a prosodic boundary at N2 when the RC was long (cf. Watson and Gibson 2004, 2005, who found more frequent prosodic boundaries after longer than after shorter constituents in unambiguous sentences).

(1a) The (unusual) plot concerns the guardian of the prince.
(1b) The prince was exiled (from the country for decades).
(1c) The (unusual) plot concerns the guardian of the prince who was exiled (from the country for decades).

Jun (2003) investigated the prosody of the sentence *Someone shot the servant of the actress who was on the balcony* in five production conditions across seven languages. In the *default reading* condition, informants read the sentence silently, answered the comprehension question *Who was on the balcony?*, and then produced the sentence twice. In all languages, the prosodic phrasings of the sentence reflected the previous RC attachment decisions. Informants who gave high attachment interpretations produced a prosodic boundary between N2 and the RC, whereas those who gave low attachment interpretations did not produce a prosodic boundary after N2. Additional informants produced a *default reading* of the sentence before giving an interpretation, and their phrasing followed the patterns described above. However, Jun does not report if they subsequently interpreted the sentence. Instead, she claims that readers' default phrasings – before or after interpretation – reflect their attachment preferences.

These production studies demonstrate that structural ambiguity and constituent length both affect lengthening of phrase-final words and insertion of boundaries. At first glance, they all seem to confirm a rather straightforward mapping between the sentential semantics and prosodic phrasing. However, the speakers in these studies always had a concrete interpretation of the sentences in mind before producing them (whether or not they were aware or made aware of the ambiguity). Thus, it remains unclear whether constituent length has a clear effect on the prosodic pattern even when speakers must read aloud the sentence while parsing its structure.

1.2 Prosody of Reading Aloud

When reading a text aloud, people generally pronounce each word as soon as it is recognized within their visual span. While reports on the prosody during reading aloud are sparse, they do suggest that readers' prosody is generated before the global syntactic structure of a sentence is determined.

Kondo and Mazuka (1996) measured eye-voice span (the distance in letters or characters between where the reader is looking and what the reader is producing) as participants were reading aloud Japanese sentences. They found that eye-voice span was only about 2.5 characters[1] regardless of the sentence's syntactic complexity. They concluded that the readers' prosody was based on limited, local syntactic information rather than a global syntactic analysis. Levin (1979) found that the eye-voice span for English is about 18 letters for good adult readers – which on average corresponds to about three or four words. Since the speaker would utter the words before grasping the overall sentential structure and meaning, the prosodic phrasing of globally ambiguous sentences may not reflect the ultimate message structure the parser achieves. Rather, it may reflect the sensible grouping of words at hand according to the local syntactic relations generated through the incremental process.

Koriat et al. (2002) also suggest that reading prosody represents local structural analysis, and precedes more complete semantic analysis. Their reading study used normal (e.g. *The fat cat with the gray stripes ran quickly*...) and nonsensical (e.g. *The sad gate with the small electricity went carefully*...) Hebrew sentences. Speakers produced prosodic patterns consistent with the local syntactic structure when reading unfamiliar sentences. In addition, semantic modifications (normal vs. nonsense sentences) did not significantly change pause patterns, suggesting that the structure of the sentence rather than its content or semantic coherence modified reading prosody. Their study thus shows that when complete semantic analysis is impeded, local structural analysis alone guides reading prosody. However, their study does not rule out that semantic analysis may modify reading prosody, e.g. in ambiguous sentences.

In sum, the above studies suggest that the prosodic patterns produced after the sentence is fully interpreted may not mirror the prosodic patterns produced while the sentence is parsed. Rather, when first reading *unfamiliar* ambiguous sentences, the local syntactic structure alone may determine readers' prosodic phrasings. Readers' subsequent comprehension judgments (after sentences are fully parsed) may not be consistent with the reading prosody produced during parsing.

1.3 Effect of Prosody on Interpreting Ambiguous Sentences

A series of research by Clifton and colleagues demonstrated the effects of boundary strengths and locations on ambiguity resolution (Carlson et al. 2001; Clifton et al. 2002, 2006). They found that the interpretation of the ambiguous structures depended on the *relative* strengths of boundaries at all relevant locations and not on one major boundary's absolute strength. Clifton

[1]Japanese orthography consists of logographic (Chinese Kanji) and syllabic (Kana/Katakana) characters.

et al. (2002) manipulated the strength of the prosodic boundary after the head noun of complex NP, as shown in (2).

(2) I met the daughter$_{[0/ip/IP]}$ of the colonel$_{ip}$ who was on the balcony.
(*0* = no boundary, *ip* = intermediate phrase boundary, IP = intonation phrase boundary, where *ip* is perceptually weaker than *IP*)

If only the local boundary at *colonel* affected the RC attachment preferences, the number of high attachment interpretations would be the same across the three conditions. However, the [0 ip] sequence elicited more high attachment choices than [ip ip], which in turn elicited more high attachment interpretations than [IP ip]. The results suggest that the interpretation of ambiguous structures is guided not by the absolute strength of a critical local boundary (as suggested by Price et al. 1991 or Marcus and Hindle 1990), but by the relative strength of the boundary in comparison to other relevant boundaries (cf. Schafer et al. 2000 above).

Clifton et al. (2006) manipulated constituent length and found that prosodic boundaries affected listeners' interpretations more when the ambiguous part of the sentence contained shorter constituents. They proposed that listeners considered prosodic boundaries to be less informative when they were produced with long constituents than with short constituents. In particular, speakers tend to produce long constituents with a preceding and following prosodic boundary. Watson and Gibson (2004, 2005) proposed that this phenomenon is related to performance: long constituents require more time for planning and recovery – time that can be gained by producing a prosodic boundary before and after the constituent. Listeners may thus take boundaries preceding or following long constituents to reflect constituent length and the performance requirements associated with it rather than syntactic structure that disambiguates the sentential meaning.

1.4 Silent Reading Studies

Silent reading studies on sentence ambiguity have focused on how RC length affects the interpretation of the RC attachment ambiguity (for English, see Bradley et al. 2003; Fernández and Bradley 1999; Hemforth et al. 2005). In these studies, participants silently read sentences and answered a comprehension question that gauged how the ambiguous constituent was attached. Across languages, longer RCs elicited more high attachment interpretations than shorter RCs. To explain these results, Fodor (2002a, 2002b) proposed the Implicit Prosody Hypothesis (IPH), given in (3).

(3) In silent reading, a default prosodic contour is projected onto the stimulus, and it may influence syntactic ambiguity resolution. Other things being equal, the

> parser favors the syntactic analysis associated with the most natural (default) prosodic contour for the construction. (2002b: 113)

> Even in reading, prosody is present. Even in silent reading, and even if prosody-marking punctuation is absent. Prosody is mentally projected by readers onto the written or printed word string. And – the crucial point – it is then treated as if it were part of the input, so it can affect syntactic ambiguity resolution in the same way as overt prosody in speech does (2002a: 83)

The implicit prosody account assumes that long RCs are more likely to be set off in their own prosodic phrase than short RCs (cf. Clifton et al. 2006; Watson and Gibson 2004, 2005). This leads to a prediction that readers are more likely to project an implicit prosodic boundary after N2 (i.e. immediately before the RC) if the following RC is long. The presence of a prosodic boundary in this location may prompt more high attachment interpretations.[2]

Since implicit prosody in silent reading cannot be measured directly, production data has often been used to support silent reading results. The implicit prosody during silent reading is assumed to match the measurable overt prosody from a production task (e.g. Fodor 1998, 2002a, 2002b; Jun 2003). However, most production studies provided interpretations of the ambiguous sentences before participants produced them (e.g. Bradley et al. 2003; Jun 2003).

1.5 The Present Study

The present study comprises two experiments on the production and comprehension of sentences with an RC attachment ambiguity, for example, *The brother*$_{N1}$ *of the bridegroom*$_{N2}$ *who swims*$_{RC}$ *was last seen on Friday night*. The production experiments investigate whether constituent length and the presence or absence of firm interpretation affects prosodic boundary placement while reading aloud such sentences. Across the two experiments, participants read aloud ambiguous sentences either before or after they chose the interpretations. In the pre-interpretation reading condition, novel sentences were read aloud on the fly and were afterwards interpreted. It is assumed that the eye-voice span during this task mimics that reported in Levin (1979), and that the prosodic patterns produced during this task better simulates the prosody during silent reading than the prosody elicited by a post-interpretation production task, where novel sentences are read silently, interpreted, and then read aloud.

[2]Many of the silent reading studies focused on local boundary strength (presence or absence of a boundary at N2) or assumed that a prosodic boundary would be projected *either* after N1 *or* after N2. Nevertheless, the results are consistent with an approach focusing on relative boundary strength since the presence of a boundary after N2 heightens the likelihood of it being the stronger of the relevant boundaries.

The length of N1 and RC was manipulated within each experiment. If constituent length modulates prosody during reading on the fly as well as during the articulation after comprehending the sentence, it would suggest that RC length shapes implicit prosody during silent reading, as the IPH proposes. If the length effect is confirmed only in the prosodic patterns produced after comprehension, it would suggest that constituent length modulates prosody only when the global sentential structure and its meaning are fully established.

The comprehension experiments tested how the prosodic patterns elicited during the two production experiments affected the interpretation of the sentences. If the relative strength of prosodic boundaries at N1 and N2 consistently reflect the message structures, a tight correspondence between the interpretation preferences and the prosodic patterns should be confirmed in the comprehension-first experiment. It remains a question of great interest whether the prosodic patterns produced before knowing the entire structure affects the subsequent interpretation of the sentence. If the interpretation of a sentence is guided by the prosodic phrasing of the articulation on the fly, prosody is confirmed to give robust cues to sentence comprehension, as suggested by Fodor (2002b). In contrast, if the interpretation of the sentence is independent of the preceding prosody, such a result would indicate that prosodic patterns generated during the initial parse are not processed as an informative cue to the comprehension of the entire sentential structure and meaning. In other words, prosody may be selectively used for sentence comprehension depending on how it is generated.

2 Experiment 1

Experiment 1 examined the prosody of ambiguous sentences unfamiliar to the reader by asking participants to read aloud sentences on a computer screen before answering a comprehension question gauging interpretation. In this task, readers were expected to build an interpretation of the sentence as they read it aloud. The experiment tested whether the length of N1 and the RC affect prosodic phrasing and whether differences in prosodic phrasing affect readers' own subsequent interpretation of the sentences.

Based on previous work (Carlson et al. 2001; Clifton et al. 2002, 2006; Fodor 1998, 2002a, 2002b; Watson and Gibson 2004, 2005), we hypothesized that long N1s and RCs should more frequently be set-off in their own prosodic phrase than short N1s and RCs. Thus, sentences with a long N1 and short RC were predicted to elicit relatively more frequent insertion of a strong boundary after N1 than after N2. Similarly, sentences with a short N1 and a long RC were predicted to elicit relatively more frequent strong boundaries after N2 than after N1. When N1 and RC are either both short or both long, frequent insertions of equal boundaries were predicted.

Table 1 Predictions for Experiment 1

Sentence type	Short N1/ long RC	Long N1 / short RC	Long N1/ long RC short N1/ short RC
⇩	⇩	⇩	⇩
Reading prosody	Boundary strength N1 < N2	Boundary strength N1 > N2	Boundary strength N1 = N2
⇩	⇩	⇩	⇩
Sentence comprehension	Fewer low attachment decisions	More low attachment decisions	Intermediate low attachment decisions

If the prosodic phrasing of participants' own speech serves as input for their subsequent sentence interpretation, a stronger prosodic boundary after N1 than after N2 should result in more frequent low attachment judgments than other reading prosodies. Also, a stronger prosodic boundary after N2 than after N1 should lead to fewer low attachment interpretations than other prosodic phrasing patterns. When the sentence is produced with equally strong boundaries at these locations, the number of low attachment interpretations should fall in the intermediate range between the other two phrasing patterns. These predictions are summarized in Table 1.

2.1 Participants

Sixteen undergraduate students at the Ohio State University participated in the study for partial course credit. They were all native speakers of, mostly Midwestern, American English. None of them reported any speech or hearing disabilities.

2.2 Materials

Twenty-four ambiguous target sentences and comprehension questions were constructed. Each sentence contained an RC modifying a complex NP, as in *The brother of the bridegroom who swims was last seen on Friday night*. Here, either the brother (high attachment) or the bridegroom (low attachment) can be the one who swims. Length of N1 (*brother*) and RC (*who swims*) were manipulated to construct four versions of each sentence, as in (4). A short N1 always consisted of two syllables, while a long N1 had four syllables. A long N1 either contained one pre-nominal modifier before the same head noun as that in the short N1 (e.g. lawyer → defense lawyer) or was a four-syllable (compound) noun semantically related to the short N1 (e.g. nanny → babysitter). A long RC contained an adjunct phrase that modifies the same verb as that of the short

RC. All short RCs consisted of two or three syllables; all long RCs had five or six syllables. The target comprehension question always gauged participants' interpretation of the ambiguity, e.g. *Who swims (like a fish)?*

(4a) Short N1 and Short RC:
The *brother* of the bridegroom *who swims* was last seen on Friday night.

(4b) Short N1 and Long RC:
The *brother* of the bridegroom *who swims like a fish* was last seen on Friday night.

(4c) Long N1 and Short RC:
The *second cousin* of the bridegroom *who swims* was last seen on Friday night.

(4d) Long NP1 and Long RC:
The *second cousin* of the bridegroom *who swims like a fish* was last seen on Friday night.

In addition, 76 filler sentences with various syntactic structures were constructed. Filler sentences were either unambiguous, contained a temporary ambiguity, or contained a global ambiguity different from the RC attachment ambiguity. Many globally ambiguous filler sentences were heavily semantically biased. None of the filler comprehension questions referred to any ambiguous part of the sentence. The most common interrogative pronoun in the filler questions was *what* (47 questions), followed by *who* (24 questions) to ensure that participants could not anticipate the interrogative pronoun and adjust their prosody in anticipation of the comprehension question. Four lists of 100 sentences were created such that each list contained only one version of the 24 target sentences.

2.3 Procedure

Participants were seated in a soundproof booth in front of a computer screen and a response box and wore a head-mounted microphone. Their productions were recorded while reading aloud 100 sentences presented one by one on the computer screen. The sentences and the comprehension questions were presented using E-Prime v.1.0 (Psychology Software Tools 2003). Participants initiated each trial by pushing a button on the response box in front of them. Upon pushing the button, a sentence appeared on the computer screen. Even though they were not given any specific instruction as to when they should begin reading, the timing of the button push and sentence production onset reveal that all participants began reading aloud each sentence immediately after it appeared on the screen, without practicing.

After reading each sentence, participants pressed a button on the response box to move on to a comprehension question about the sentence. Three response options were presented underneath the question. Participants responded by pushing the button that corresponded to their answer choice. Target questions always inquired about who performed the action in the RC, e.g. *Who swims (like a fish)?* The response options always corresponded to NP1 (e.g. *the brother / the second cousin*), NP2 (e.g. *the bridegroom*), and “I don’t know”. Participants’ response to the comprehension question prompted the beginning of the next trial. Before the experiment, participants had the opportunity to practice the task and ask any questions about the instructions given to them.

2.4 ToBI Coding

The productions were coded using the ToBI (Tones and Break Indices; see Beckman and Hirschberg 1993; Beckman and Ayers 1997; Silverman et al. 1992) intonation transcription system for Standard American English (SAE). The ToBI system for SAE is based on the autosegmental metrical theory of intonation originally proposed by Pierrehumbert (1980) (see Beckman, Hirschberg and Shattuck-Hufnagel (2007) for a history of the ToBI framework). The theory assumes two levels of phrasal boundaries for English: perceptually weaker intermediate phrase (ip) boundaries and perceptually stronger intonational phrase (IP) boundaries. Only the ToBI break indices were used to code the productions.[3] Every word was given a break index that indicated the strength of prosodic juncture at its right edge. The following indices were used:

(5) ToBI Break Indices:
1 = word boundary
2 = hesitation
3 = intermediate phrase (ip) boundary (perceptually weaker boundary)
4 = Intonational Phrase (IP) boundary (perceptually stronger boundary)

The first author of the paper, who is trained in the ToBI system, transcribed all productions. A second coder unaffiliated with the project transcribed a small

[3]In a production study that used the same task as Experiment 1, Bergmann, Armstrong, and Maday (2008) compared the production of sentences like *Someone shot the servant of the actress who was standing on the balcony* in English and Spanish. They coded prosodic boundary strength and prosodic boundary type at the verb, N1 and N2. They found that boundary strength in English different at the two sentence locations N1 and N2 with more IP boundaries at N2 than N1. Boundary type at N1 and N2, on the other hand, were comparable: Almost all ip boundaries at N1 and N2 had a H- phrase accent and over 80% of IP boundaries had the patterns H-L% (floor-holding pattern) or L-H% (continuation rise). These boundary types all convey that the speaker is not done speaking (cf. Pierrehumbert & Hirschberg 1990). Prosodic boundaries at the end of the sentence were overwhelmingly L-L%, indicating finality. The boundaries at N1 and N2 thus seem to reflect the fact that the sentence is not over. We therefore focus on boundary strength in our analysis.

subset of productions. Intertranscriber reliability of the coders is 89% for all words and 79% for the critical words (N1 and N2) (cf. intertranscriber reliabilities in Syrdal and McGory 2000 and Dilley et al. 2006). For data-internal consistency, the data reported here are based on the first coder's transcriptions.

2.5 Phonetics

Even though ToBI labeling yielded relatively stable results across labelers, it is a subjective measure. One of the most consistent phonetic cues to a prosodic boundary is pre-boundary lengthening (e.g. Lehiste 1973). We therefore measured duration of N1 and N2 to support the ToBI coding of the data (Fig. 1). The words that received higher break indices tended to have longer durations: The mean and median value of the distribution rises for each higher break index. Pairwise comparisons (Tukey contrasts) revealed that at N1 words with no boundary were reliably shorter than words with an ip boundary ($z = 2.9, p < 0.05$) or IP boundary ($z = 7.0, p < 0.001$). Words with an ip boundary were marginally shorter than words with an IP boundary ($z = 2.2, p = 0.07$). At N2 words with an IP boundary were reliably longer than words with either an ip boundary ($z = 7.8, p < 0.001$) or no boundary ($z = 7.4, p < 0.001$).

The results of a correlation analysis are found in Table 2. The analysis showed only a weak positive correlation of prosodic boundary strength and word duration. Only between 12% and 14% of the variability in the data can be accounted for by the strength of prosodic boundary. However, word duration correlates rather strongly with the words' numbers of segments, which accounts for 73% of the variability in duration of N1 and 51% of the variability of N2. Finally, word

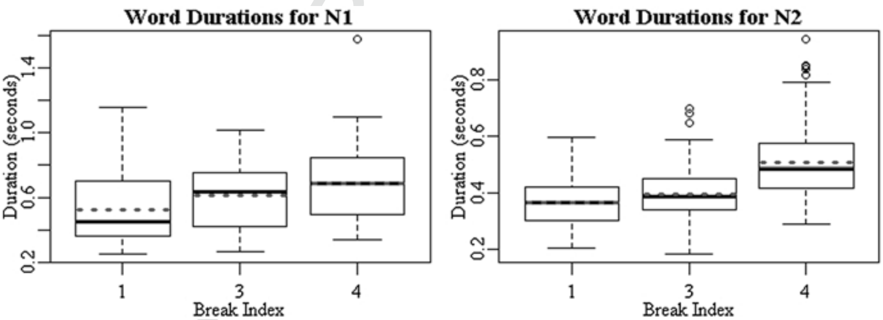


Fig. 1 Boxplots showing the distribution of duration measurements across boundary types at N1 and N2. For each break index, the solid black line shows the median and the gray dotted line shows the mean of the distribution. The box below the median shows the second quartile, the one above the median shows the third quartile. The bottom whisker shows the first quartile, the top whisker shows the fourth quartile. The dots represent outliers. The position of the median within the box indicates the skew of the distribution (skewed right if the median is near the bottom of the box and vice versa)

Table 2 Correlation analysis of word durations from Experiment 1

	Prosodic boundary strength	Number of segments	Word frequency
Word duration N1	$r = 0.34; r^2 = 0.12$	$r = 0.86; r^2 = 0.73$	Not available*
Word duration N2	$r = 0.38; r^2 = 0.14$	$r = 0.71; r^2 = 0.51$	$r = -0.48; r^2 = 0.23$

*Due to English spelling conventions, where many compound nouns are spelled as two separate words (e.g. *defense lawyer*), it was not possible to obtain frequency counts for many N1s.

frequency of N2 (using Kučera-Francis written frequency counts, Kučera and Francis 1967) showed a weak negative correlation with word duration, accounting for 23% of the variability of N2 durations. Together, strength of prosodic boundary, number of segments, and word frequency account for at least 85% of the variability in word duration. We suggest that much of the remaining variability can be accounted for by individual differences in reading speed.

2.6 *Results*

2.6.1 Prosodic Phrasing

The experiment elicited a possible total of 384 utterances. However, participants failed to correctly produce a total of six utterances. Six further utterances were excluded because they contained a disfluency at a critical location. The prosodic patterns of the remaining 372 utterances were categorized into three groups according to the location and relative strength of the boundaries. Sentences with a stronger prosodic boundary at N1 than at N2 (*IP*/*ip* or *ip*/*0*) were categorized as Stronger Break Follows (SBF) N1, and those with a stronger boundary at N2 than at N1 (*ip*/*IP* or *0*/*ip*) were labeled SBF N2. Sentences with equally strong prosodic boundaries at both sentence locations (*IP*/*IP*, *ip*/*ip* or *0*/*0*) were categorized as Equally Strong Breaks (ESB). These global prosodic patterns, which capture the relative boundary strengths at the relevant sentence locations, should be related to how the sentences are interpreted (cf. Carlson et al. 2001; Clifton et al. 2002). Figure 2 compares the occurrences of these prosodic patterns across the four sentence types. Table 3 shows the results from a multinomial logistic regression predicting the location of the stronger boundary from N1 and RC length. The results reveal that sentences with Short N1s (as opposed to Long N1s) were produced reliably more often with ESB ($p < .01$) and SBF N2 ($p < .001$) than with SBF N1. No such differences were found for the RC length manipulation. The results suggest that the length of N1, but not the length of the RC, affected the relative strength of prosodic boundaries at N1 and N2 when people were reading the sentences on the fly.

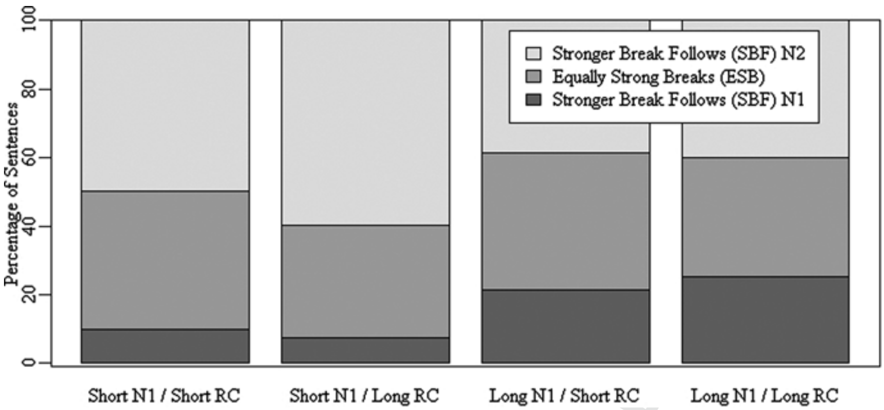


Fig. 2 Global prosodies of interest elicited by each sentence type. The x-axis shows the sentence types, and the y-axis shows the percentage of sentences produced with each prosodic pattern of interest

Table 3 Multinomial logistic regression for Figure 2: Coefficients of predicted factors

Factors	SBF N1 vs. ESB			SBF N1 vs. SBF N2			ESB vs. SBF N2		
	Est.	t	$p <$	Est.	t	$p <$	Est.	t	$p <$
Long N1 vs. Short N1	0.965929	2.8537	.01	1.323249	4.0142	.001	0.35732	1.5540	n.s.
Long RC vs. Short RC	0.220067	0.7017	n.s.	−0.069763	−0.2275	n.s.	−0.28983	−1.2625	n.s.

For all *X vs. Y*, *X* represents the baseline and *Y* represents the alternative. $N = 372$; log-likelihood = −368.57; McFadden $R^2 = 0.025694$; likelihood ratio test: $\chi^2 = 19.44$, $p < 0.001$.

Next, we coded the local boundaries at N1 and N2 according to their strength (no boundary, *ip* boundary or *IP* boundary) across the different length conditions to see if longer constituents are more likely to be set off in their own prosodic phrase (as suggested by Fodor 1998, 2002a, 2002b, Watson and Gibson 2004, 2005). Figure 3a compares the distribution of boundary strengths at N1 for the Short N1 and Long N1 sentences, whereas Fig. 3b compares the distribution of boundary strengths at N2 for the Short RC and Long RC sentences. Tables 4a and 4b show the results from multinomial logistic regressions predicting local boundary strength at N1 from N1 length and local boundary strength at N2 from RC length. Sentences with a long N1 were more likely to be produced with a prosodic boundary (either ip ($p < .05$) or IP ($p < .001$) as opposed to no boundary) after N1 than sentences with a short N1. This suggests that long N1s were more frequently produced as their own prosodic phrase than short N1s, as suggested by Fodor (1998, 2002a, 2002b) and Watson and Gibson (2004, 2005). However, the same relation between constituent length and prosodic phrasing did not hold for RCs: The insertion of a boundary after N2 was equally frequent across the sentences with a long RC and those with a short RC, i.e., a long RC did not increase the likelihood of a

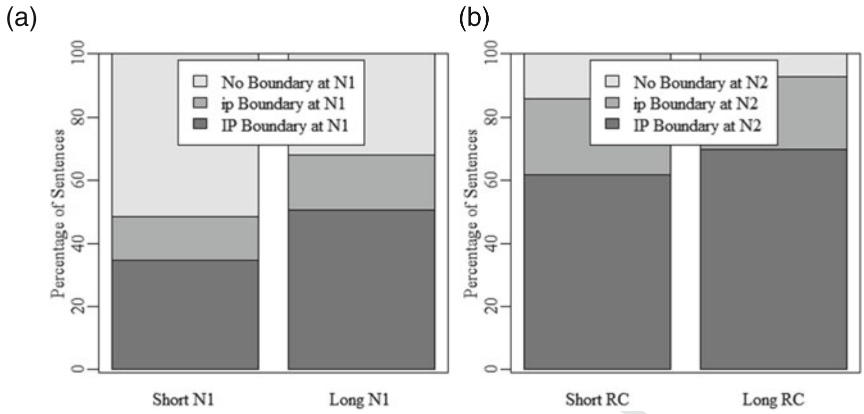


Fig. 3 (**a**) Number and type of prosodic boundaries at N1 from sentences with Short and Long N1s (**b**) Number and type of prosodic boundaries at N2 from sentences with Short and Long RCs

preceding prosodic juncture. Rather, participants produced reliably more IP boundaries than ip ($p < .001$) and more ip boundaries than no boundary ($p < .01$) at N2, regardless of RC length. In addition, the presence or absence of a prosodic boundary after N1 did not affect boundary placement after N2. Rather, RCs were mostly preceded by a prosodic boundary regardless of their length and regardless of whether readers had just produced a boundary at N1.

The present results do not support the hypothesis that long constituents are unconditionally more likely to be produced as their own prosodic phrase. While the length of N1 affected the location and relative strength of prosodic boundaries for the entire sentence, the length of the RC did not. In addition, only the

Table 4a Multinomial logistic regression for Figure 3a: Coefficients of predicted factors

Factors	No boundary vs. *ip* boundary at N1			No boundary vs. *IP* boundary at N1			*ip* boundary vs. *IP* boundary at N1		
	Est.	*t*	*p* <	Est.	*t*	*p* <	Est.	*t*	*p* <
Long N1 vs. short N1	−0.73716	−2.3609	.05	−0.85453	−3.6998	.001	−0.11736	−0.3779	n.s.

For all *X vs. Y*, *X* represents the baseline and *Y* represents the alternative. $N = 372$; log-likelihood = −371.08; McFadden $R^2 = 0.019965$; likelihood ratio test: $\chi^2 = 15.119$, $p < 0.001$.

Table 4b Multinomial logistic regression for Fig. 3b: Coefficients of predicted factors

Factors	No boundary vs. *ip* boundary at N2			No boundary vs. *IP* boundary at N2			*ip* boundary vs. *IP* boundary at N2		
	Est.	*t*	*p*<	Est.	*t*	*p* <	Est.	*t*	p <
Long RC vs. Short RC	−0.31449	−0.8128	n.s.	−0.52035	−1.4983	n.s.	−0.20585	−0.8267	n.s.

For all *X vs. Y*, *X* represents the baseline and *Y* represents the alternative. $N = 372$; log-likelihood = −317.67; McFadden $R^2 = 0.0040533$; likelihood ratio test: $\chi^2 = 2.5857$, $p = 0.27449$.

length of N1 predicted the likelihood of a prosodic boundary after N1, but the length of the RC did not predict the likelihood of a prosodic boundary after N2. This suggests that prosodic phrasing of a sentence may be modulated by the length of only certain types of constituents.

One potential factor behind this discrepancy is the semantic and syntactic independence of the morphemes and words added to N1 and RC. While long N1s were nouns or compound nouns with merely more syllables than the nouns of the short N1s, long RCs always had additional words that formed a constituent on their own. As a result, there may have been more opportunities to produce a prosodic boundary within the RC (e.g. who swims // like a fish) than within N1 (e.g. second // cousin). This additional boundary between the verb and the adjunct phrase may have weakened the effect of the overall length of the RC. However, an additional multinomial logistic regression revealed that this was not the case. The Long RC sentences that were produced with a boundary after the verb were not less likely to be produced with a prosodic boundary after N2 than those without a boundary after the verb (all p-values > 0.1).

A more plausible reason for the asymmetric effect of the constituent length is the differences in the availability of length information during production. Note that a prosodic boundary after N1 was inserted *after* participants uttered the entire noun phrase, while a boundary after N2 was inserted *before* the RC for which the length was manipulated. Thus, participants may have had different degrees of certainty about the length of the constituents at those two locations. The null effect of the RC length on the likelihood of a boundary after N2 may reflect that participants had not established the length of the upcoming RC as certainly as they had established the length of N1. As readers produced N2, the entire RC was likely not always within their eye-voice span. It is plausible that participants could anticipate an RC following N2 due to the relative pronoun ‘who’, but they could not produce a prosodic boundary that would reflect the size of the following RC at this juncture.

2.6.2 Sentence Comprehension

A total of 372 responses to the comprehension questions were coded according to the participants’ interpretation judgments: high-attachment (*the brother / second cousin swims*), low-attachment (*the bridegroom swims*), and *don’t know*. Only six sentences received *don’t know* responses: They were excluded from the analysis. The distributions of interpretation judgments were compared across the three prosodic patterns to examine the relation between the prosodic phrasing and the following interpretation of the sentences (Fig. 4). Results from a multinomial logistic regression are found in Table 5. The results indicate that interpretation choices were not guided by the prosodic patterns participants produced, i.e., a stronger boundary after N1 than after N2 did not elicit more low attachment responses than a stronger boundary after N2 than after N1. Instead, all prosodic patterns elicited comparable proportions of low attachment and high attachment responses, and low attachment was the dominant interpretation after any of the three prosodic renditions. These results were

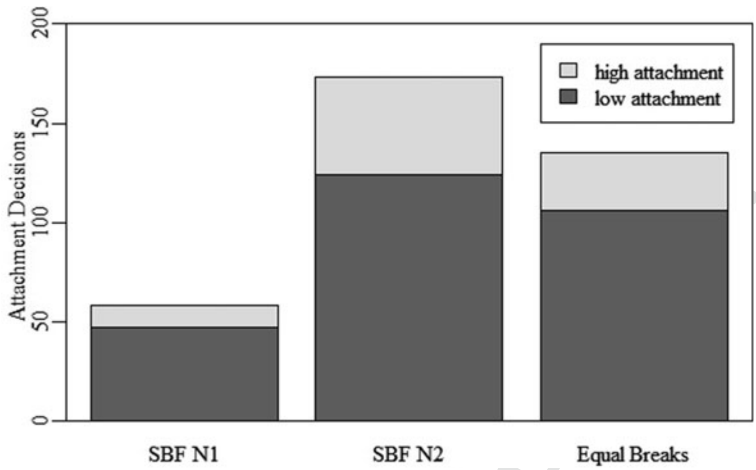


Fig. 4 Sentence interpretations by global prosodic patterns. The x-axis shows the prosodic patterns produced; the y-axis shows the sentence interpretations

Table 5 Multinomial logistic regression for Fig. 4: Coefficients of predicted factors

Factors	Low attachment vs. High attachment		
	Est.	t	$p <$
ESB vs. SBF N1	0.15611	0.3951	n.s.
ESB vs. SBF N2	−0.36768	−1.3666	n.s.
SBF N1 vs. SBF N2	−0.52379	−1.3966	n.s.

For all *X vs. Y*, *X* represents the baseline and *Y* represents the alternative. $N = 372$; log-likelihood $= -201.51$; McFadden $R^2 = 0.0074332$; likelihood ratio test: $\chi^2 = 3.0182$, $p = 0.22111$.

somewhat surprising since they conflict with Jun's (2003) production data and since relative boundary strength exhibited much clearer effects on attachment preferences in Carlson et al. (2001) and Clifton et al. (2002).

3 Experiment 2

Experiment 2 examined the prosody produced after readers became familiar with the sentences. In this experiment, readers silently read the sentence and answered the comprehension question before they read it aloud. The experiment tests whether the length of N1 and the RC affect the interpretations of the sentences and whether the interpretations in turn affect readers' subsequent prosodic phrasing.

Experiment 1 showed that only the length of N1, but not of the RC, affected prosody when unfamiliar sentences were read aloud. If we assume that the implicit prosody produced during silent reading (i.e. in this experiment) is similar to the overt prosody in reading aloud on the fly (i.e. in Experiment 1), only the length of N1 should affect readers' implicit prosody. We would then expect more implicit prosodic boundaries after long N1s than after short N1s and frequent implicit prosodic boundaries after N2 regardless of RC length. According to the result from Experiment 1, RC length should not affect interpretation decisions.

Alternatively, implicit reading prosody may differ from overt reading prosody. In particular, silent reading may require fewer processing resources than reading aloud and may allow for more preview of the upcoming structure. If so, readers may have a better idea of the length of the RC when N2 is parsed in silent reading. As a result, even RC length may matter for the implicit prosodic phrasing. Thus, in silent reading, both long N1s and RCs may be set-off in their own prosodic phrase more often than short N1s and RCs. If this implicit prosody guides the interpretation, which in turn feeds the overt prosody, the phrasing patterns produced after silent reading and after the comprehension response may better mirror the implicit prosody than does the explicit prosody during reading on the fly.

If RC length modulates the implicit reading prosody that guides the interpretation, sentences with a short N1 and a long RC are expected to have a stronger implicit boundary after N2 than after N1 and elicit fewer low attachment interpretations than other sentence types. Sentences with a long N1 and a short RC are expected to have a stronger implicit boundary after N1 than after N2, and thus should elicit more low attachment interpretations than other sentence types. Sentences with a short N1 and a short RC or with a long N1 and a long RC should elicit an intermediate number of low attachment judgments. If sentence interpretation affects subsequent reading prosody, low attachment interpretations should induce more productions of a stronger prosodic boundary after N1 than after N2 than high attachment interpretations. High attachment interpretations are expected to lead to more productions of a stronger boundary after N2 than after N1 than low attachment interpretations. These predictions are summarized in Table 6.

Table 6 Predictions for Experiment 2

Sentence Type	Short N1 / long RC	Long N1 / short RC	Long N1/ long RC short N1/ short RC
⇩	⇩	⇩	⇩
Silent Reading	Implicit boundary strength N1 < N2	Implicit boundary strength N1 > N2	Implicit boundary strength N1 = N2
⇩	⇩	⇩	⇩
Sentence Comprehension	Fewer low attachment decisions	More low attachment decisions	Intermediate low attachment decisions
⇩	⇩	⇩	⇩
Reading Prosody	More: boundary strength N1 < N2	More: boundary strength N1 > N2	All Patterns: N1 < N2; N1 > N2; N1 = N2

3.1 Participants

Sixteen undergraduate students from Ohio State University participated in the study for partial course credit. They were all native speakers of, mostly Midwestern, American English. None of them reported any speech or hearing disabilities. No participant in Experiment 2 had participated in Experiment 1.

3.2 Materials

The same 24 target items and 76 filler sentences as in Experiment 1 were used for Experiment 2. The comprehension questions were also identical to those used in Experiment 1.

3.3 Procedure

The procedure for Experiment 2 differed from Experiment 1 mainly in the order of events. During each trial, participants first read each sentence silently. They were instructed to make sure they understood the meaning of the sentence, and there was no time limit for reading the sentence. After silently reading the sentence, participants pushed a button on the response box to move on to the comprehension question with the response options. After responding to the question, the sentence reappeared on the screen and participants were recorded reading the sentence aloud. They were instructed to concentrate on getting the meaning of the sentence across.

3.4 ToBI Coding

The productions were coded as in Experiment 1.

3.5 Phonetics

The durations of N1 and N2 were measured for each relevant break index (Fig. 5). The words that received higher break indices tended to have longer durations: The mean and median value of the distribution rises for each higher break index. Pairwise comparisons (Tukey contrasts) revealed that at N1 words with an IP boundary were reliably longer than words with either an ip boundary ($z = 5.3$, $p < 0.001$) or no boundary ($z = 7.8$, $p < 0.001$) and that at N2 words with an IP boundary were reliably longer than words with either an ip boundary ($z = 5.1$, $p < 0.001$) or no boundary ($z = 8.5$, $p < 0.001$).

Table 7 shows the results of a correlation analysis. Again we find only a weak positive correlation of prosodic boundary strength and word duration. Between

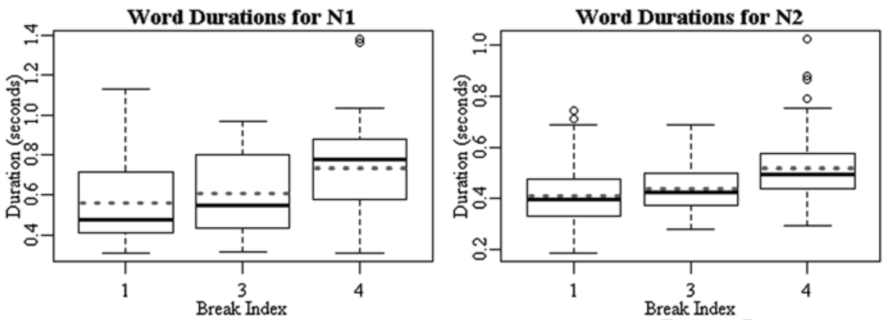


Fig. 5 Boxplots showing the distribution of duration measurements across boundary types at N1 and N2. For each break index, the solid black line shows the median and the gray dotted line shows the mean of the distribution. The box below the median shows the second quartile, the one above the median shows the third quartile. The bottom whisker shows the first quartile, the top whisker shows the fourth quartile. The dots represent outliers. The position of the median within the box indicates the skew of the distribution (skewed right if the median is near the bottom of the box and vice versa)

Table 7 Correlation analysis of word durations from Experiment 2

	Prosodic boundary Strength	Number of segments	Word frequency
Word Duration N1	$r = 0.34$; $r^2 = 0.12$	$r = 0.87$; $r^2 = 0.75$	not available*
Word Duration N2	$r = 0.39$; $r^2 = 0.15$	$r = 0.68$; $r^2 = 046$	$r = -0.47$; $r^2 = 0.22$

*Due to English spelling conventions, where many compound nouns are spelled as two separate words (e.g. *defense lawyer*), it was again not possible to obtain frequency counts for many N1s.

12% and 15% of the variability in the data can be accounted for by the strength of prosodic boundary. Again number of segments correlates fairly strongly with word duration, such that the strength of prosodic boundary, number of segments, and word frequency (for N2) together account for 83% or more of the variability in word duration.

3.6 Results

3.6.1 Sentence Comprehension

Only eight of the 384 interpretation judgments elicited *don't know* responses. These were therefore excluded from the analysis, leaving a total of 376 interpretation judgments. No sentence was excluded for disfluencies at critical locations. Figure 6 compares the distribution of interpretation judgments across the four types of sentences. While the overall dominance of low attachment interpretations

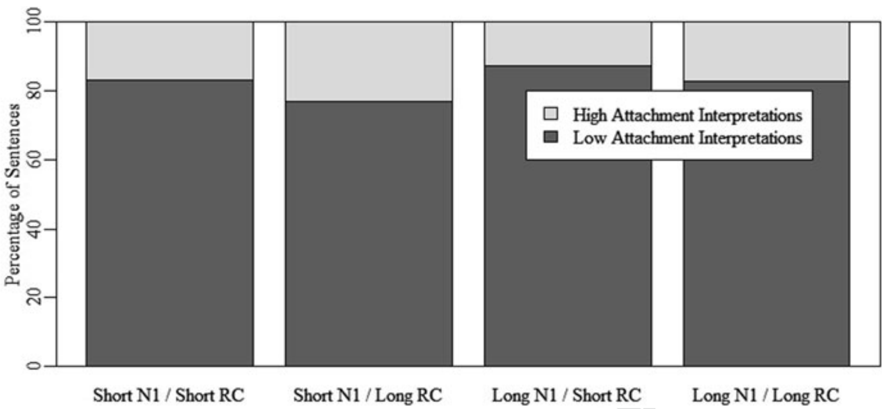


Fig. 6 Interpretations of each sentence type. The x-axis shows the sentence types, and the y-axis shows the percentage of sentences given a high attachment vs. low attachment interpretation

Table 8 Multinomial logistic regression for Fig. 6: Coefficients of predicted factors

Factors	Low attachment vs. High attachment		
	Est.	t	$p <$
Long N1 vs. Short N1	0.35678	1.3002	n.s.
Long RC vs. Short RC	−0.36970	−1.3474	n.s.

For all *X vs. Y*, *X* represents the baseline and *Y* represents the alternative. $N = 372$; log-likelihood $= -172.9$; McFadden $R^2 = 0.010162$; likelihood ratio test: $\chi^2 = 3.55, p = 0.16949$.

was confirmed, the four sentence types elicited the predicted differences in interpretation patterns only numerically. Sentences with short N1 s and long RCs elicited the fewest number of low attachment interpretations (76.8%). Sentences with long N1s and short RCs elicited the highest number of low attachment interpretations (87.2%). Finally, sentences for which both N1 and RC were either short or long elicited an intermediate number of low attachment interpretations (83% and 82.8%, respectively). However, these differences were not statistically reliable (see Table 8). Therefore, the effect of constituent length on the attachment preference does not seem to be as robust as previously reported (e.g. Bradley et al. 2003; Fernández and Bradley 1999; Hemforth et al. 2005).

3.6.2 Sentence Interpretation and Prosodic Patterns

Recall that Experiment 1 showed no link between the interpretation choices and the global prosodic patterns that readers previously produced. In Experiment 2, readers' interpretation choices affected the global prosodic patterns they subsequently produced. Figure 7 shows the distribution of prosodic patterns produced for the sentences of each interpretation choice. The results of a

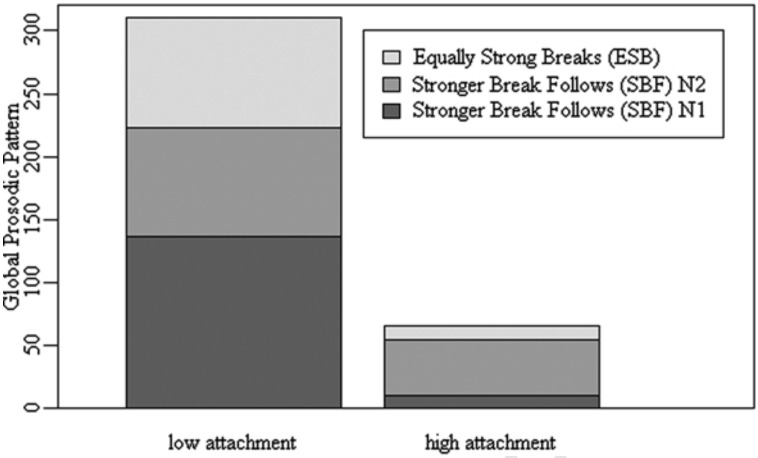


Fig. 7 Sentence interpretations by global prosodic patterns. The x-axis shows the prosodic patterns produced; the y-axis shows the sentence interpretations

multinomial logistic regression show that low attachment interpretations (compared to high attachment interpretations) more often elicited productions with the stronger boundary following N1 than with the stronger boundary following N2 (Table 9). They also elicited more productions with equally strong boundaries than with the stronger boundary following N2 ($p < 0.001$). That is, participants were reliably less likely to produce sentences with the stronger boundary following N2 when they had given the sentence a low attachment interpretation than when they had given it a high attachment interpretation.

Contrary to our prediction, low attachment interpretations elicited a number of productions with equally strong boundaries after N1 and N2 (28.1%) and even with the stronger boundary following N2 (27.7%). In other words, low attachment interpretations did not elicit productions with the stronger boundary after N1 (only 44.2%) as often as expected. Recall that Fig. 4 shows that without having been forced to provide an interpretation prior to production, readers strongly disprefer this prosodic pattern (15.8% of total productions). This general dispreference may explain the lower than expected number of productions with the stronger boundary following N1.

Table 9 Multinomial logistic regression for Fig. 7: Coefficients of predicted factors

Factors	SBF N1 vs. ESB			SBF N1 vs. SBF N2			ESB vs. SBF N2		
	Est.	*t*	p <	Est.	*t*	*p* <	Est.	*t*	*p* <
high att. vs. low att.	−0.63639	−1.4155	n.s.	−1.94724	−5.1737	.001	−1.310844	−3.6471	.001

For all *X vs. Y*, *X* represents the baseline and *Y* represents the alternative. $N = 372$; log-likelihood = −389.86; McFadden $R^2 = 0.045011$; likelihood ratio test: $\chi^2 = 36.751, p < 0.001$.

3.6.3 Constituent Length and Prosodic Patterns

The results further reveal that the global prosodic patterns depended on the lengths of the constituents. Figure 8 shows the distribution of the prosodic patterns elicited by each one of the four sentence types. As in Experiment 1, sentences with Short N1s elicited fewer productions with the stronger prosodic boundary after N1 than with the stronger boundary after N2 than sentences with Long N1s. Unlike Experiment 1, RC length affected the global prosodic pattern. Sentences with Long RCs reliably elicited more productions with the stronger prosodic boundary after N2 than equally strong boundaries or the stronger boundary following N1 than sentences with Short RCs (Table 10). This suggests that when the lengths of the constituents are known before production and when readers try to convey their interpretations, constituent length overall affects the prosodic patterns of the sentences.

Contrary to predictions, sentences with Short N1 and RC and Long N1 and RC did not elicit an 'intermediate' production pattern. Rather, sentences with Short N1 and RC elicited a pattern similar to sentences with Long N1 and Short RC, and sentences with Long N1 and RC elicited a pattern similar to sentences with Short N1 and Long RC.

Next, the distributions of the boundary types were grouped according to the length of N1 (Fig. 9a) and RC (Fig. 9b). Statistical analysis confirms the effect of N1 length found Experiment 1, i.e., sentences with a long N1 were more likely to be produced with an IP boundary after N1 (compared to no boundary ($p <$ 0.001) or an ip boundary ($p < 0.001$)) than sentences with a short N1 (Table 11a). Long N1s are thus more frequently produced as their own prosodic phrases than Short N1s.

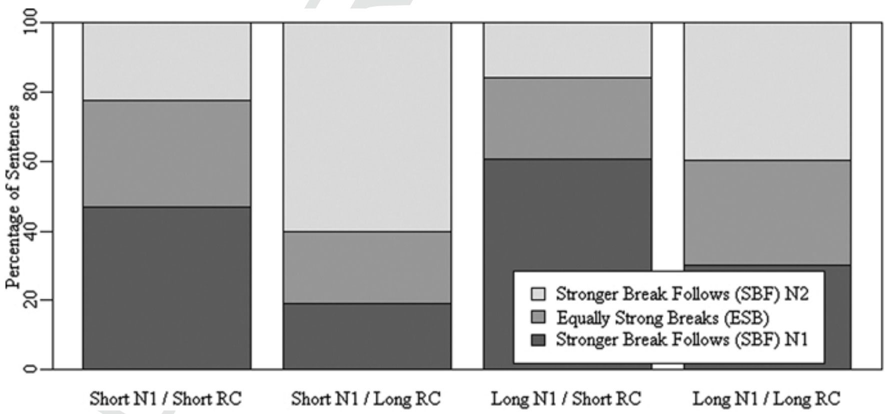


Fig. 8 Global prosodies of interest elicited by each sentence type. The x-axis shows the sentence types, and the y-axis shows the number of sentences produced with each prosodic pattern of interest

Table 10 Multinomial logistic regression for Fig. 8: Coefficients of predicted factors

Factors	SBF N1 vs. ESB			SBF N1 vs. SBF N2			ESB vs. SBF N2		
	Est.	t	$p <$	Est.	t	$p <$	Est.	t	$p <$
Long N1 vs. Short N1	0.335394	1.2664	n.s.	0.816579	3.1001	.01	0.48118	1.7478	n.s.
Long RC vs. Short RC	−0.744738	−2.7615	.01	−1.792098	−6.6434	.001	−1.04736	−3.7040	.001

For all X *vs.* Y, X represents the baseline and Y represents the alternative. $N = 372$; log-likelihood = −379.3; McFadden $R^2 = 0.070872$; likelihood ratio test: $\chi^2 = 57.865, p < 0.001$.

Figure 9b demonstrates a pattern unlike that found in Experiment 1. Contrary to any predictions, sentences with Long RCs were less likely to be produced with a prosodic boundary after N2. Instead, sentences with Short RCs were more often produced with either an ip boundary ($p < 0.001$) or an IP boundary than no boundary ($p < 0.001$) than sentences with Long RCs (Table 11b).

Notice also that regardless of RC length, productions in Experiment 2 had much fewer prosodic boundaries after N2 (Fig. 9b) than in the productions from Experiment 1 (Fig. 3b). This suggests that readers' overall low attachment preference for these sentences is reflected in the absence of a boundary at N2 when the sentences were previously interpreted (Experiment 2), but not when the sentences were not previously interpreted (Experiment 1).

Finally, unlike Experiment 1, boundary placement at N2 was modulated by the presence or absence of a prosodic boundary at N1 (Fig. 10). Readers were less likely to produce an ip or IP boundary at N2 (as opposed to no boundary) if they had produced an ip or IP boundary at N1 (as opposed to no boundary) (Table 12).

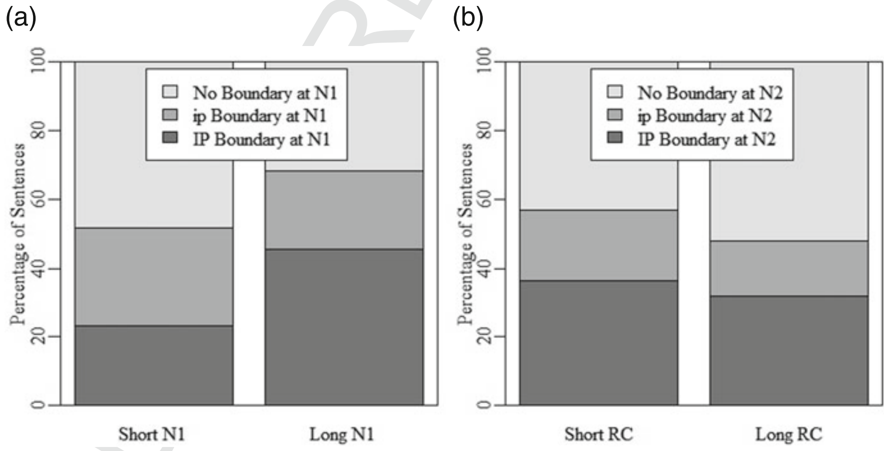


Fig. 9 (**a**) Number and kind of prosodic boundaries at N1 from sentences with Short and Long N1s (**b**) Number and kind of prosodic boundaries at N2 from sentences with Short and Long RCs

Table 11a Multinomial logistic regression for Fig. 9a: Coefficients of predicted factors

Factors	No boundary vs. *ip* boundary at N1			No boundary vs. *IP* boundary at N1			*ip* boundary vs. *IP* boundary at N1		
	Est.	t	$p <$	Est.	t	$p <$	Est.	t	$p <$
Long N1 vs. Short N1	–0.11857	−0.4442	n.s.	−1.04767	−4.2338	.001	−0.92910	−3.3327	.001

For all X *vs.* Y, X represents the baseline and Y represents the alternative. $N = 372$; log-likelihood = −395.3; McFadden $R^2 = 0.025662$; likelihood ratio test: $\chi^2 = 20.823$, $p < 0.001$.

Table 11b Multinomial logistic regression for Fig. 9b: Coefficients of predicted factors

Factors	No boundary vs. *ip* boundary at N2			No boundary vs. *IP* boundary at N2			*ip* boundary vs. *IP* boundary at N2		
	Est.	t	$p <$	Est.	t	$p <$	Est.	t	$p <$
Long RC vs. Short RC	−1.53248	−5.0597	.001	−1.62779	−6.4913	.001	−0.09531	−0.2990	n.s.

For all X *vs.* Y, X represents the baseline and Y represents the alternative. $N = 372$; log-likelihood = −360.16; McFadden $R^2 = 0.071194$; likelihood ratio test: $\chi^2 = 55.214$, $p < 0.001$.

The results of Experiment 2 do not support the hypothesis that long constituents in general are more likely to be produced as their own prosodic phrase: while the length of N1 affected the local placement of a prosodic boundary after N1, it was the presence or absence of a boundary after N1, rather than the length of the RC, that affected the presence or absence of a boundary after N2. However, the results show that both the length of N1 and RC affect the relative strength of boundaries within a sentence.

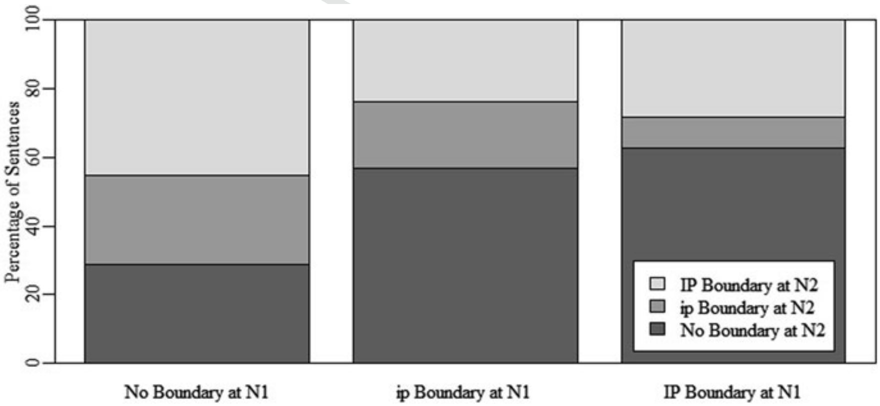


Fig. 10 Boundary placement at N2 as a function of boundary placement at N1. The x-axis shows boundary types at N1; the y-axis shows boundary types at N2

Table 12 Multinomial logistic regression for Fig. 10: Coefficients of predicted factors

Factors	No boundary vs. *ip* boundary at N2			No boundary vs. *IP* boundary at N2			*ip* boundary vs. *IP* boundary at N2		
	Est.	*t*	*p* <	Est.	*t*	*p* <	Est.	*t*	*p* <
no vs. *ip* at N1	−0.95929	−2.7376	.01	−1.32917	−4.1711	.001	−0.36987	−0.9845	n.s.
no vs. *IP* at N2	−1.80118	−4.7482	.001	−1.24572	−4.5056	.001	0.55547	1.4319	n.s.
ip vs. *IP* at N2	–0.841892	−2.0422	.05	0.083448	0.2593	n.s.	0.92534	2.0127	.05

For all *X vs. Y*, *X* represents the baseline and *Y* represents the alternative. $N = 372$; log-likelihood $= -367.05$; McFadden $R^2 = 0.053442$; likelihood ratio test: $\chi^2 = 41.447$, $p < 0.001$.

4 Discussion

Combining the simple tasks to read aloud sentences and answer comprehension questions, the present study investigated how constituent length and the familiarity with the target sentences modified explicit reading prosody. Experiment 1 revealed that in the case of reading aloud unfamiliar sentences, constituent length affected the likelihood of boundary insertion only when information about the size of the constituent was fully established. In Experiment 2, the knowledge of constituent length affected the relative strength of prosodic boundaries more than it affected the presence or absence of local boundaries. In addition, the location of a preceding prosodic boundary and readers' sentence interpretations both modulated reading prosody.

These results suggest different mechanisms for reading familiar and unfamiliar sentences. For unfamiliar sentences, local syntactic cues that become available within the span of preview seem to primarily guide the prosodic phrasing, as suggested by Kondo and Mazuka (1996) and Koriat et al. (2002). In Experiment 1, only the length of N1, but not of the RC, predicted the presence or the absence of boundaries and the relative strength of boundaries across the two locations (i.e., after N1/N2). Thus, a longer constituent was produced with a boundary only afterward, but not beforehand. We suspect that this asymmetry emerged because the certainty of locally available syntactic structure differed across the two locations. When readers fixated the first content word of the sentence, the size of N1 was probably fully noticed, as the following preposition 'of' must have been reviewed within the para-foveal window – which is known to span 14–15 character spaces (for a review, see Rayner 1975, 1992, 1998). When the reader detected that the subject noun phrase contained a following PP, the longer letter string or another content word that required the articulation of additional two syllables might have prompted the insertion of a boundary after N1, especially because the size of following modifier PP was probably not yet available. Assuming that the eye-voice span of our participants was up to 18 character spaces (Levin 1979) and that the frequent functional words such as articles and copulas were often not fixated (see the review on past text-reading studies in Rayner and Liversedge 2004), readers' eyes must have already been viewing N2 for sentences with short N1s, whereas they were fixating 'of' or the region beyond it for sentences with long N1s when their voicing started. Therefore, we suspect that the insertion of a post-N1 boundary was possibly planned even before the first word was uttered.

On the contrary, the full length of the RC may not have been available to the readers when they were about to finish articulating N2. Readers were most likely previewing the relative pronoun *who* and the following verb as they started producing N2. Readers therefore knew that N2 was followed by an RC, but the word following the verb that determined the length of the RC may not have been fully processed to establish the constituent size while planning the prosodic phrasing upon articulating N2. Even if the word following the verb was within the preview window, the RC pronoun *who* may have warned readers about the potential complexity of the upcoming constituent: It is plausible that readers overwhelmingly placed a boundary after N2 in preparation for a potentially complex and semantically dense constituent.

When the sentence had been comprehended and had become familiar, both the lengths of constituents and the global syntactic structure that was based on the interpretation seemed to shape the prosody (cf. Jun 2003, where participants' interpretation also affected the subsequent production of the prosodic boundaries). In Experiment 2, the length of N1 affected both the local boundary placement and the relative strength of boundaries across the two locations. Readers again separated a Long N1 from the following PP, and did so more often when the following RC had been interpreted as low-attached and thus was not separated from N2. Overall, the preference for low attachment interpretations led to much fewer boundaries after N2, supporting Carlson and Frazier's (2002) rational speaker hypothesis. In addition, the presence and the strength of the prosodic boundary after N1 affected the presence and the strength of a boundary after N2 more than the length of the RC. These results together suggest that the knowledge of the lengths of all constituents and the interpretation-driven syntactic structure prompted readers to use prosodic boundaries more informatively and to avoid close boundary placement within a sentence. Thus, higher familiarity with the sentence seems to allow not only better control over the informative phrasing of the word string, but also a better rhythmic adjustment of the utterance.

In Experiment 2, the ShortN1/LongRC and LongN1/ShortRC sentences induced the expected productions of stronger breaks after N2 and after N1, respectively. However, the two sentence types with either both Short or both Long N1 and RC were not produced with similar prosodic patterns. Instead, sentences with Short N1 and Short RC were likely to be produced with the stronger boundary after N1, while sentences with Long N1 and Long RC were often produced with the stronger boundary after N2. Thus, the sentences with short constituents were produced as the low-attached structure, which in Jun's (2003) study was the default reading for English. When constituents were longer, however, the largest prosodic boundary was more frequently inserted before the beginning of the complex RC. Although it may not be directly motivated by the attachment preference, a prosodic juncture following and preceding a large constituent may have been required for rhythmically felicitous speech planning.

In addition to the effect of familiarity on the overt prosodic patterns, the present study demonstrated the limits of prosodic effects on the interpretation of ambiguous sentences. In contrast to past findings of listeners' sensitivity to relative prosodic boundary strength (Clifton et al. 2002, 2006), Experiment 1 showed that the prosodic patterns produced while simultaneously parsing the sentences did not guide the subsequent interpretation of the sentence. As discussed above, the prosodic phrasings while reading aloud unfamiliar sentences were guided by local syntactic information rather than global semantic information. When readers were asked to provide interpretations of the sentences after reading them on the fly, they could access the complete syntactic structure and reconstruct the semantic relations among the constituents. During this reconstruction of the message, readers seemed to ignore the prosody they previously produced. Thus, prosodic phrasing may serve as an input to the sentence interpretation only when the listener believes that the prosody cues the syntactic/semantic structure of the sentence.

Finally, an account must be provided for the lack of an effect of constituent length on sentence interpretation. Across the two experiments, we found a robust low-attachment preference that was not modulated by the length of the RC. Such results do not conform to those of previous studies (Bradley et al. 2003; Fernández and Bradley 1999; Hemforth et al. 2005). One possible reason for the absence of a length effect in the present study is insufficient difference in length across the sentence types. Long RCs in previous studies were often over five syllables longer than their short counterparts. In our study, long RCs were only between three and five syllables longer than the corresponding short RCs. Hemforth et al. (2006) propose that longer RCs are usually also more informative than shorter RCs (e.g. *who swims* vs. *who swims like a fish*), and require more costly semantic processing. They further argue that the presence of a semantically more demanding long RC can be justified if it modifies the central element of the proposition, such as the *head* of the subject noun phrase (i.e., N1). According to this "information load account", more high-attachment responses may have been observed if our Long RCs were semantically heavier.

5 Conclusions

The present study exhibited how familiarity with the sentence and semantic structure affects the prosodic patterns of structurally ambiguous sentences. Although the analyses focused on the strengths of boundaries at the two RC attachment sites and their relations to sentence interpretations, the prosodic cues to the sentence structure are not limited to the size of boundaries. For example, Schafer et al. (2000) found that listeners successfully determined the intended syntax of the ambiguous sentences in spontaneous speech even when the relative strengths of the boundaries at relevant locations did not support this interpretation. Thus, future studies need to examine how listeners make use of other prosodic cues such as phrasal accents, boundary tones, the types and distribution of pitch accents, rhythm, and microprosody to achieve the comprehension of sentence structure intended by the speaker.

Using the utterances produced in the present study, we are currently investigating whether listeners are sensitive to the speaker's familiarity with the sentences they produced. If naïve listeners relied solely on the boundary strength to achieve the interpretation of the sentence, utterances of the same prosodic patterns would lead to identical interpretations regardless of the speaker's familiarity. Preliminary data suggests that listeners find the unfamiliar speakers' prosody less reliable than that of the familiar speakers. We plan to further investigate whether the perceived degree of speaker's certainty about the sentential message relates to perceived fluency, as well as what prosodic cues predict the perception of speaker's certainty or fluency.

References

Altmann, Gerry T. M. 1998. Ambiguity in sentence processing. *Trends in Cognitive Sciences* 2: 146–152.

Beckman, Mary E. and Julia Hirschberg. 1993. *The ToBI Annotation Conventions*. Unpublished manuscript.

Beckman, Mary E. and Gayle M. Ayers. 1997. *Guidelines for ToBI Labeling*. Unpublished manuscript.

Beckman, Mary E., Julia Hirschberg and Stefanie Shattuck-Hufnagel. 2007. The original ToBI System and the evolution of the ToBI framework. In Sun-Ah Jun (ed.) *Prosodic Typology: The Phonology of Intonation and Phrasing*, 9–54. Oxford: Oxford University Press.

Bergmann, Anouschka, Meghan Armstrong and Kristine Maday. 2008. Relative clause attachment in English and Spanish: A production study. *Proceedings of the Conference on Speech Prosody 2008*, Campinas, Brazil, 505–508.

Bradley, Dianne, Eva M. Fernández and Dianne Taylor. 2003. *Prosodic Weight versus Information Load in the Relative Clause Attachment Ambiguity*. Paper presented at the 16th CUNY Conference on Human Sentence Processing. Cambridge, MA.

Carlson, Katy, Charles Clifton, Jr. and Lyn Frazier. 2001. Prosodic boundaries in adjunct attachment. *Journal of Memory and Language* 45: 58–81.

Carreiras, Manuel and Charles Clifton, Jr. 1993. Relative clause interpretation preferences in Spanish and English. *Language and Speech* 36: 353–372.

Carreiras, Manuel and Charles Clifton, Jr. 1999. Another word on parsing relative clauses: Eye-tracking evidence from Spanish and English. *Memory and Cognition* 27: 826–833.

Clifton, Charles Jr. and Lyn Frazier. 1996. *Construal*. Cambridge, MA: MIT Press.

Clifton, Charles Jr. Katy Carlson and Lyn Frazier. 2002. Informative prosodic boundaries. *Language and Speech* 45: 87–114.

Clifton, Charles Jr. Katy Carlson and Lyn Frazier. 2006. Tracking the what and why of speakers' choices: Prosodic boundaries and the length of constituents. *Psychonomic Bulletin and Review* 13: 854–861.

Cuetos, Fernando and Don C. Mitchell. 1988. Cross-linguistic differences in parsing: Restrictions on the use of the Late Closure strategy in Spanish. *Cognition* 30: 73–105.

Dilley, Laura, Mara Breen, Edward Gibson, Marti Bolivar and John Kraemer. 2006. A comparison of inter-coder reliability for two systems of prosodic transcriptions: RaP (Rhythm and Pitch) and ToBI (Tones and Break Indices). *INTERSPEECH-2006*, paper 1619-Mon2A3O.6.

Fernández, Eva M. and Dianne Bradley. 1999. *Length Effects in the Attachment of Relative Clauses in English*. Poster presented at the 12th Annual CUNY Conference on Human Sentence Processing. New York, NY

Fodor, Janet Dean. 1998. Learning to parse? *Journal of Psycholinguistic Research* 27: 285–319.

Fodor, Janet Dean. 2002a. Prosodic disambiguation in silent reading. *Proceedings of the North East Linguistic Society (NELS)* 32: 113–132. Amherst, MA.

Fodor, Janet Dean. 2002b. Psycholinguistics cannot escape prosody. *Proceedings of Speech Prosody*, 83–88. Aix-en-Provence, France.

Hemforth, Barbara, Susana Fernández, Charles Clifton, Jr., Lyn Frazier and Lars Konieczny. 2005. *Relative clause attachment in German, English, Spanish \and French: Effects of position and length*. Manuscript.

Hemforth, Barbara, Caterina Petrone, Mariapaola D'Imperio, Joël Pynte, Saveria Colonna and Lars Konieczny. 2006. Length effects in PP-attachment: Prosodic balancing or balancing information load. *Proceedings of the Annual Conference of the Cognitive Science Society*. Vancouver, Canada.

Jun, Sun-Ah. 2003. Prosodic phrasing and attachment preferences. *Journal of Psycholinguistic Research* 32: 219–249.

Kimball, John. 1973. Seven principles of surface structure parsing in natural language. *Cognition* 2: 15–47.

Kondo, Tadahisa and Reiko Mazuka. 1996. Prosodic planning while reading aloud: On-ine examination of Japanese sentences. *Journal of Psycholinguistic Research* 25: 357–381.

Koriat, Asher, Seth N. Greenberg and Hamutal Kreiner. 2002. The extraction of structure during reading: Evidence from reading prosody. *Memory and Cognition* 30: 270–280.

Kučera, Henry and W. Nelson Francis. 1967. *Computational Analysis of Present-Day American English*. Providence: Brown University Press.

Lehiste, Ilse. 1973. Rhythmic and syntactic units in production and perception. *Journal of the Acoustical Society of America* 54: 1228–1234.

Levin, Harry. 1979. *The Eye-Voice Span*. Cambridge, MA: MIT Press.

Marcus, Mitchell and Donald Hindle. 1990. Description theory and intonation boundaries. In Gerry T. M. Altmann (ed.) *Cognitive models of speech processing: Psycholinguistic and Computational Perspectives*, 483–512. Cambridge, MA: MIT Press.

Miyamoto, Edson T. 2008. *Relative Clause Attachment*. http://etm4rc.googlepages.com/table.html. Accessed 21 April 2010.

Pierrehumbert, Janet B. 1980. *The Phonology and Phonetics of English Intonation*. Ph.D. Dissertation. MIT, Cambridge, MA.

Pierrehumbert, Janet B. and Julia Hirschberg. 1990. The meaning of intonational contours in the interpretation of discourse. In Philip R. Cohen, Jerry Morgan and Martha E. Pollack, (eds.) *Intentions in Communication*, 271–311. Cambridge, MA: MIT Press.

Price, Patti J.; Mari Ostendorf and Stefanie Shattuck-Hufnagel. 1991. The use of prosody in syntactic disambiguation. *Journal of the Acoustical Society of America* 90: 2956–2970.

Psychology Software Tools, Inc.. 2003. *E-Prime*. [computer software].

Rayner, Keith. 1975. The perceptual span and peripheral cues in reading. *Cognitive Psychology* 7: 65–81.

Rayner, Keith. 1992. *Eye Movements and Visual Cognition: Scene Perception and Reading*. Springer Series in Neuropsychology. New York, NY: Springer-Verlag.

Rayner, Keith. 1998. Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin* 124: 372–422.

Rayner, Keith and Simon P. Liversedge. 2004. Visual and linguistic processing during eye fixations in reading. In Fernanda Ferreira and John M. Henderson (eds.), *The Interface of*

Language, Vision and Action: Eye Movements and the Visual World, 59–104. New York: Psychology Press.

Schafer, Amy J., Shari R. Speer, Paul Warren and S. David White. 2000. Intonational disambiguation in sentence production and comprehension. *Journal of Psycholinguistic Research* 29: 169–182.

Silverman, Kim, Mary E. Beckman, John Pitrelli, Mari Ostendorf, Colin Wightman, Patti J. Price, Janet Pierrehumbert and Julia Hirschberg. 1992. ToBI: A standard for labeling English prosody. *Proceedings of the 1992 International Conference of Spoken Language Processing* 2: 867–870.

Syrdal, Ann K and Julie McGory. 2000. Inter-transcriber reliability of ToBI prosodic labeling. *Proceedings of the International Conference on Spoken Language Processing*. Beijing, China, 235–238.

Watson, Duane and Edward Gibson. 2004. The relationship between intonational phrasing and syntactic structure in language production. *Language and Cognitive Processes* 19: 713–755.

Watson, Duane and Edward Gibson. 2005. Intonational phrasing and constituency in language production and comprehension. *Studia Linguistica* 59: 279–300.

Semantically-Independent but Contextually-Dependent Interpretation of Contrastive Accent

Kiwako Ito and Shari R. Speer

1 Introduction

Successful communication requires agreement on what has been and is being talked about. As the status of discourse information is constantly updated across the time-course of a conversation, interlocutors continuously distinguish topics and propositions, identify the most relevant discourse entity of the moment, and allocate processing resources to achieve a sufficient interpretation of the message. At times, it becomes important and beneficial to both parties of a conversation to explicitly discern what needs to be foreground among all the candidate discourse entities at hand. Prosody is a powerful tool in this act.

The present investigation concerns the effect of the presence of a L + H* accent on pre-nominal adjectives. The core function of a distinctive salient accent was described decades ago by Bolinger (1961), who identified the tight link between accentuation and the act of singling out a specific item from a larger but limited set, and distinguished such accentual prominences from those that do not evoke contrast from the set. The present research aims to assess the effect of contrast-evoking pitch prominence, which may or may not be conditioned by the semantics of the accented words. For the purpose of annotation, we adopt the conventions of the ToBI system (Tone and Break Indices: Beckman and Ayers 1997; Beckman, et al. 2005), and attempt to advance our understanding of the pragmatic function of L + H* beyond the proposal by Pierrehumbert and Hirschberg (1990). According to their view, entities uttered with L + H* are distinct from other accents in that 'the accented item – and not some alternative related item – should be mutually believed (p. 296).' This statement echoes Bolinger's description of accent function in implicating that the alternatives are evoked together with the accented item. However, it remains unclear precisely how the set of alternatives is generated and updated as a

K. Ito (✉)
Department of Linguistics, Ohio State University, Columbus, Ohio, US
e-mail: ito@ling.ohio-state.edu

S. Frota et al. (eds.), *Prosodic Categories: Production, Perception and Comprehension*, Studies in Natural Language and Linguistic Theory 82,
DOI 10.1007/978-94-007-0137-3_4,

discourse transpires, and thus how contrastive accentual prominence affects the online interpretation of reference in spoken sentences.

Previous eye-movement monitoring studies demonstrated the effect of prominent pitch accent on a contrastive adjective in English during visual search (Ito and Speer 2008). This work expanded on a previous study by Dahan et al. (2002), who captured the non-anaphoric interpretation of accented nouns in participants' eye movements during a simple screen-based object manipulation task. In the Dahan et al. experiments, participants were presented with cohort-paired objects (e.g., *candle* and *candy*) and a couple of distractors (e.g., *necklace* and *pear*), and followed sequential instructions to move those objects around a set of immovable geometric shapes (e.g., *Put the candy above the triangle*. *Now, put the CANDLE/candle above the square*.) When an object was repeated from the first to the second instruction (e.g., *candle*-*candle*), the fixation proportions to the cohort competitor (e.g., *candy*) were higher during the second mention when it had accentual prominence than when it did not. When the object was altered (e.g., *candy*-*candle*), lack of prominence in the second instruction increased looks to the previously mentioned object as compared to the prominent mention. These results demonstrate the immediate use of prosody and imply that accentual prominence directs attention to previously unmentioned objects in the visual display. However, because the prosodic manipulation occurred directly on the noun, which provided the segmental label for the referent and thus strictly constrained the activation of candidates, these experiments did not allow us to observe how prosody alone would guide referential resolution.

Pre-nominal adjectives are often used to more precisely specify a referent, restricting the set of candidate referents even before the noun's segmental information can single out the target referent (e.g., 'Didn't you recently get a new burgundy cashmere sweater?'). Thus, manipulating the accentual prominence of pre-nominal adjectives permits the examination of how prosody can constrain the set of alternatives for the referent. Sedivy et al. (1999) compared the effect of a prominent accent on a pre-nominal modifier to a condition without such an accent. They found no effect of accent on referent identification. This may have been due to their use of a highly contrastive visual layout with a long pre-utterance visual exposure and a minimal number of objects (e.g., *pink comb, yellow comb, yellow bowl, metal knife*). These factors may have allowed participants' eye movements to reach their ceiling speed, i.e., their detection of the target was indistinguishably fast regardless of the presence of prominence on the adjective. Ito and Speer (2008) manipulated the accentual prominence of adjectives in instructions for a visual search among objects in a more complex layout. Participants followed sequential directions to decorate holiday trees (e.g., *First, hang a blue ball*. *Next, hang a green drum*.). Each target noun phrase was composed of a color adjective and an object noun, and its prosodic pattern varied among [H* !H*], [L + H* no-accent], and [H* L + H*]. Ornaments of various colors were sorted by object type across a grid, and participants were asked to hang ornaments one by one until they finished decorating each of the four trees. In Experiment 1, felicitous L + H* on the adjectives in sequences with repeated nouns (e.g., *green drum* → *BLUE*$_{\mathrm{L+H*}}$ *drum*$_{\text{no-acc}}$) facilitated fixations to targets as compared to infelicitous renditions (e.g., *blue*$_{\mathrm{H*}}$ *DRUM*$_{\mathrm{L+H*}}$). Infelicitous prominence on the repeated adjective (e.g., *blue ball* → *BLUE* $_{\mathrm{L+H*}}$ *drum*$_{\text{no-acc}}$) led to slower fixations to the target than the felicitous renditions (e.g., *blue* $_{\mathrm{H*}}$ *DRUM* $_{\mathrm{L+H*}}$). In Experiment 2, the facilitative effect of L + H* on the adjective was confirmed in comparison to neutral H*!H* renditions (e.g., *blue*$_{\mathrm{H*}}$ *drum*$_{\mathrm{!H*}}$). Furthermore, infelicitous L + H* on the adjective in sequences where the noun did not repeat (e.g., *red onion* → *GREEN*$_{\mathrm{L+H*}}$ *drum*$_{\text{no-acc}}$) led to incorrect anticipatory fixations to the previously mentioned noun (here, *onion*), which began toward the end of the adjective and lasted until about 300 ms into the target noun.

The results of Ito and Speer (2008) indicate that a prominent L + H* accent on a pre-nominal adjective evokes a set of alternatives via the most salient referent – such as the most recently mentioned object type in a discourse. The activation of the contrastive alternatives is robust, as shown in the initial incorrect fixations to the previous ornament set even as the segmental information for a different ornament noun was being presented - a 'garden-path' effect in eye movements. Similar results have been obtained in German (Weber et al. 2006) and Japanese (Ito et al. 2009; ms.), suggesting that prosodic prominence on the pre-nominal modifier may have a general effect of restricting the candidates for the upcoming noun across languages. It is important to note that the relatively large number of objects used for the visual search was advantageous for detecting prosody-based activation of an alternative set of referents in the discourse structure. Unlike in Dahan et al., Weber et al., and Sedivy et al., where the targets were displayed with a single competitor, participants in Ito and Speer were exposed to many objects (40–52 per grid) across multiple instructions (24–26 per tree). Since ornaments of the same color were distributed across multiple cells, the incorrect fixations to the just-mentioned object set (e.g., looks to the 'onion' cell upon hearing '*GREEN drum*') confirmed the selective activation of the contrast-evoked alternative (e.g., green onion) out of the other visually accessible candidates (e.g., green bell, green ball, and green stocking, etc.). That is, anticipatory fixations were detected not because the referents contrasted visually with recently mentioned objects, but because the adjectives' prosodic prominence was instantly linked to the salience of a recently activated reference through the dynamic discourse updating mechanism.

While how pre-nominal L + H* activates the alternative referents was rather straightforwardly demonstrated in Ito and Speer (2008), there remains a possibility that findings benefited from the design of their visual layout. In both the locally contrastive (e.g., blue drum → GREEN drum) and non-contrastive (e.g., red onion → GREEN drum) sequences, anticipatory effects were found in the eye movements returning to the cell of the previously mentioned ornament. Saccades to known locations are certainly easier to program and more precisely executed (Henderson and Ferreira 2004), so those fairly swift fixation patterns might have resulted from the fact that the participants' eyes were returning to recently-visited cells. Thus, the anticipatory effect of L + H* must be re-attested with a layout where the set of alternatives are located separately from the contrast counterparts. The present experiments are therefore designed to overcome this potential design artifact of Ito and Speer (2008).

More importantly than this methodological refinement, the present study examines the interaction between the contrast-evoking effect of L + H* and the semantics of pre-nominal adjectives. Although the function of prosodic prominence seems to be fairly general and robust, its effect may be largely subsidiary to the restrictive function of word meaning. That is, although prosodic prominence produced anticipatory looks to a contrastive target when the contrast was determined by a color adjective within a reference set containing other colors, such effects might not hold when the contrastive interpretation is generated by the semantics of a pre-nominal adjective that is inherently contrastive. Note that all of the abovementioned work (Sedivy et al. 1999; Weber et al. 2006; Ito and Speer 2008; Ito et al. 2009; ms) tested the effect of prosodic prominence with color adjectives. Color adjectives in those experimental contexts are often classified as *intersective* adjectives, as the referent sets denoted by the adjectives (e.g., red things) intersect with another set denoted by nouns (e.g., drums) to yield the restricted set of entities (e.g., red drums). Adjectives such as *French*, *healthy* and *wooden* belong to this category, as the interpretations of the properties denoted by these adjectives do not depend on the semantics of the combined adjectives or nouns. Adjectives such as *big, good*, and *old* are instead called *subsective* adjectives, as the truth conditions of the adjectives are evaluated with respect to the referent set denoted by the combined nouns. For example, a twenty-year-old car could be considered old, whereas a twenty-year-old temple could not really be called an ‘old temple.’ Rather than specify independent, absolute properties of the entities, *subsective* pre-nominal adjectives determine the subset of the entities referred to by the following noun. (For more formal and theoretical treatment of the adjective distinctions, see Chierchia and McConnell-Ginet 2000; Chierchia and Turner 1988; Kamp 1975; Kamp and Partee 1995; Oltean 2007, inter alia.) Since a subsective adjective requires a relative interpretation, the presence of such an adjective may automatically evoke a notion of contrast, regardless of the prosody with which it is pronounced (e.g., Use of ‘big’ in ‘*Give me a big cup’* may connote the presence of smaller cups).

In the present experiments, the effect of L + H* is tested with two sets of adjectives – colors and sizes-with a tree decoration task akin to that in Ito and Speer (2008). The ornaments are sorted by the object type AND the color/size, such that the prosody-driven anticipation (e.g., red drum → GREEN/LARGE . . .) can be observed in the eye-movements to a non-repeated location. In the context of the experimental search space, the *intersective* color adjectives (red, green and yellow) convey discrete values in hue, whereas the *subsective* size adjectives (small, medium and large) require comparisons among the set of visually accessible entities. With the color adjectives, we should observe both facilitative and garden-path effects analogous to those reported in Ito and Speer (2008) provided that L + H* invariably evokes robust anticipation for a contrast. If the new layout-which may increase the difficulty in visual search-substantially delays the fixation timings as compared to those of our previous study, we would have to reconsider how prosodic processing interacts with the time course of scene perception. As for the size adjectives, at least two outcomes are possible. On one hand, if the presence of a size adjective evokes a stronger notion of contrast than a color adjective *and* if the contrast-evoking effect of L + H* is robust independently of the adjective's semantics, the anticipatory and the garden-path effects may then become even more evident with size adjectives than with color adjectives. On the other hand, if the effect of accentuation is only complementary to the semantics of the adjective, the contrastive prominence of L + H* may be redundant in the presence of the contrast-evoking size adjectives, leading to no difference in fixation timing. If L + H* is found to have no effect with the size adjectives, such a result would suggest that lexical semantics modulate prosodic processing, i.e., lexical semantics are – at least partially-processed before prosody is processed, falsifying our former claim that prosody is processed on a par with segmental information.

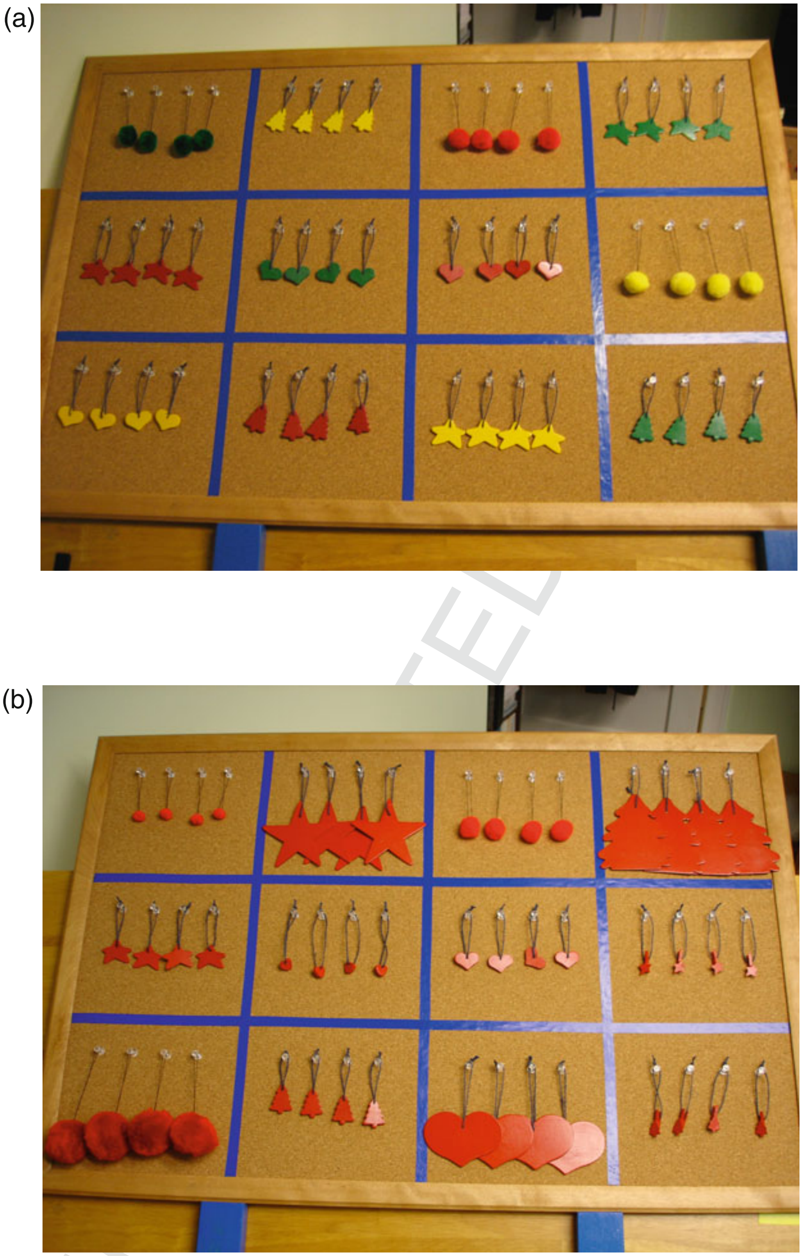
2 Experiments

2.1 Participants

Thirty-seven and forty undergraduate students at Ohio State University participated in the Color and the Size experiments, respectively. Participants earned partial course credit for their participation. No student participated in both experiments. Data from seven Color- and four Size-participants were excluded from the analyses due to eye calibration failure, frequent track loss or because the participant's native language was not English.

2.2 Design and Materials

Four types of ornaments (balls, trees, stars, and hearts) were prepared in three sizes (small, medium, and large) and in three colors (red, green and yellow). For the Color Experiment, three display boards were created on which ornaments were of comparable size (i.e., small, medium, and large sets). Each board had a total of 48 ornaments displayed in 12 cells, such that each cell contained four identical ornaments, with different colors distributed across cells. (Fig. 1a shows the 'medium' set). The size differences across the three boards created variety in the search task, and served to distract participants from the experimental manipulation. For the Size Experiment, the same sets of ornaments were displayed by color (i.e. red, green, and yellow sets). Again, each of the three boards contained 48 ornaments, the four identical ornaments in each cell of the same color, with different sizes distributed across cells (Fig. 1b shows the 'red' set). In both



This figure will be printed in b/w

Fig. 1 Example image of ornament board (1a) Color experiment: medium ornament set; (1b) Size experiment: red ornament set experiments, the locations of the ornaments were randomly assigned except that the two center cells of the middle row were always occupied by hearts – the foil ornaments that did not participate in the critical trials. Those fillers were placed in the central cells to roughly equate the distance from the center point of the visual field to each critical ornament cell, and also to reduce the occurrences of uninterpretable fixations during the para-foveal detection of the critical ornaments, which would be more frequent than desired were the targets clustered in the central region of the visual field.

The auditory instructions for the tree decoration were recorded with a trained female phonetician who maintained her overall pitch range and speech rate within and across conditions, and also across experiments. In both experiments, each of the three trees was decorated with 26 ornaments, for a total of 156 instructions (26 ornaments $\times$ 3 trees $\times$ 2 experiments). These were recorded along with a set of utterances specifying other aspects of the decoration process (e.g., whether to proceed vertically-from the top to the bottom or in the opposite direction- or horizontally-from the left to the right or in the opposite direction- and how many rows of ornaments, how many ornaments in each row, etc.). Decoration directions were included on the recording script for each tree, and thus the speaker produced the instructions in the actual order of decoration. In both experiments, critical instructions were either in a locally contrastive sequence where the noun was repeated from the previous instruction (e.g., green ball $\rightarrow$ yellow ball; medium tree $\rightarrow$ small tree), or in a locally non-contrastive sequence where neither the adjective nor the noun was repeated (e.g., red tree $\rightarrow$ yellow star; large tree $\rightarrow$ medium ball). For both types of sequence, the adjective-noun combination was produced with either [L + H* no-accent L-L%] or [H* !H* L-L%] accentual pattern.

For both experiments, the combinations of the adjectives and the nouns were counterbalanced such that no particular combination of color/size and ornament type would be associated with a specific accentual pattern. Within a tree, each of the three adjectives was mentioned 8 or 9 times, and each of the four ornament nouns was mentioned 6 or 7 times. The critical 9 combinations of the adjective and the noun, i.e., 3 adjectives (color or size) $\times$ 3 nouns (ball/star/tree), appeared once in each of the four conditions. The referential noun phrase was always produced at the end of a naturalistic utterance such as '*O.K., to its right, hang the <u>yellow ball</u>*'. Care was taken such that the phrases or discourse markers that were used to increase naturalness (e.g., *O.K., Next, Moving to the left*, etc.) did not signal any particular informational status for the upcoming referent (e.g., L + H* was never used for the phrases preceding the noun phrase. For the null effect of accentual prominence on a discourse marker, see Metusalem and Ito 2008). To control the discourse context across conditions, the instruction immediately preceding the critical instruction was produced with [H* !H* L-L%] across conditions. All instructions were randomized and submitted to ToBI-annotation by two trained labelers who were blind to the manipulations of the discourse structure. Each critical instruction was re-recorded until the two labelers agreed that the pitch shape reflected the intended accentual pattern.

Table 1 Mean duration and mean F0 values of the target stimuli

	Adj		Noun	
Experiment sequence accent	Dur (ms)	F0 (Hz)	Dur (ms)	F0 (Hz)
Color Contrastive [L + H* no-acc] red star → <u>YELLOW</u> star	330	299	489	148
Color Contrastive [H* !H*] red star → <u>yellow</u> star	332	207	549	164
Paired t Stat (df = 8)	.09	11.17***	2^	4.42**
Color Non-Contrastive [L + H* no-acc] red star → <u>YELLOW</u> star	320	300	491	150
Color Non-Contrastive [H* !H*] red star → <u>yellow</u> star	316	208	558	163
Paired t Stat (df = 8)	.18	12.39***	1.51	8.26***
Size Contrastive [L + H* no-acc] medium star → <u>LARGE</u> star	483	289	493	145
Color Contrastive [H* !H*] medium star → <u>large</u> star	484	213	570	167
Paired t Stat (df = 8)	.08	7.87***	2.32*	9.92***
Color Non-Contrastive [L + H* no-acc] red star → <u>YELLOW</u> star	480	283	512	148
Color Non-Contrastive [H* !H*] red star → <u>yellow</u> star	480	214	578	166
Paired t Stat (df = 8)	.004	7.33***	1.56	25.69***

***$p < .001$, **$p < .01$, *$p < .05$, ^$p < .1$

In addition to the ToBI annotation, duration and F0 measures confirmed acoustic dissimilarities between [L + H* no-accent L-L%] and [H* !H* L-L%]. Table 1 summarizes the average duration and F0 of the adjectives and nouns in each condition of the two experiments. Although L + H* accentuation on the adjectives yielded much higher F0 than H*, such pitch distinction was not accompanied by a difference in duration in either experiment. The nouns following L + H* adjectives were attenuated in both duration and F0 as compared to those following H* adjectives (The difference in F0 contours is exemplified in Fig. 2a and 2b).

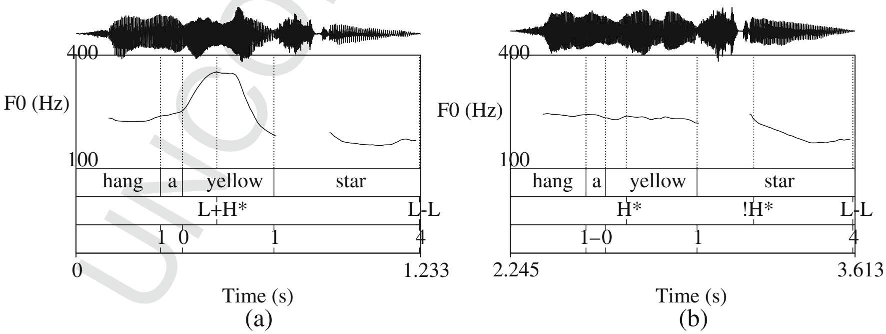


Fig. 2 Example ToBI transcriptions of target noun phrases in the Color experiment: *Hang a YELLOW star* with [L + H* no-accent]: and (2b) *Hang a yellow star* with [H* !H*]

2.3 *Procedure*

Participants were seated in front of a drafting table with the top tilted at 35 degrees to support the ornament board. They wore a lightweight headgear (ASL 6000 Head Mounted Optics) fitted with an eye-camera and a magnetic transmitter that functioned to correct the measured eye positions for head movement. Before decorating the trees, each participant's left eye was calibrated using the 9-point calibration system. After successful calibration, participants followed pre-recorded instructions played through a set of speakers placed behind the drawing table. After each instruction, participants choose an ornament from the board and placed it on a small tree located to their right, and then faced back to the board and waited for the next instruction. The x and y coordinates of eye-fixations on the ornament board were recorded at 60 Hz using ASL Eye-Trac 6 data-collection system. The experimenter monitored the participant's eye locations and body orientations captured by a ceiling-mounted camera, and pressed a key to play each instruction when the participant faced back to the board after hanging the correct ornament. The sequencing and order of the tree decoration instructions were varied across the three trees to reduce the predictability in the task and ensure participants' attention throughout the experiment. Each experimental session lasted from forty-five minutes to an hour.

3 Results

In what follows, we present the fixation data in logit terms (log odds). For graphing, the fixation data was coded as either 1 (on) or 0 (off) on a given area of interest (AOI) and the logit was calculated based on the ratio between 1 s and 0 s for each time point (See Barr 2008; Jaeger 2008; and Johnson 2008 for discussion of binomial data coding and logit transformation). A total of four (out of 270) trials in the Color, and twenty-one (out of 324) trials from the Size experiment were excluded from analysis because the participant made a mistake in selecting the ornament, or requested the replay of an instruction as s/he failed to pay attention to it, or the monocle that reflected the near-infrared light was displaced during the critical instruction. In the figures, the logit data is time-aligned at the boundaries between the adjective and the noun, where the segmental information of the noun started to provide the critical information for identifying the target referent. The fixation likelihood was plotted backwards from this point for the adjective and the preceding part of the instruction, and forwards for the target noun and the following region up to 1000 ms post noun onset.

3.1 *L + H* Facilitates Visual Search*

With a visual layout designed to compel a search for contrastive objects in novel locations, the facilitative effect of L + H* was confirmed in both the Color and the Size experiments. Figures 3a and 3b compare fixation likelihood to the target cells between two locally contrastive sequences in each experiment. In both experiments, the likelihood to fixate on the contrastive target increased more swiftly when the pre-nominal adjective carried L + H* (e.g., Color: red star → *YELLOW* star; Size: medium star → *LARGE* star) than when it was produced with H* (e.g., *yellow/large* star). Both figures show steep rises in the looks to the target cell beginning toward the end of the adjectives, indicating that participants started their search for the upcoming referent before they heard the noun. This search was evidently much faster in the L + H* trials than in the H* trials from an early stage, as shown in the steeper rise of the fixation likelihood for L + H* in both 3a and 3b.

To test the effect of accentual prominence on the fixation to the target, we applied mixed effects logistic regression models that specified both subject and item as random effects (Jaeger 2008) to the data coded into 300 ms windows.[1] Following Barr (2008; Barr & Frank, 2009), we first plotted the mean logit function across the two conditions to determine the time region where the fixation likelihood should be modeled, and confirmed that the fixation increase was initiated at around -300 ms and reached its ceiling range at around 900 ms *regardless of the prosodic manipulation* in both experiments.[2] This determined an analysis window from -300 to 900 ms. In order to reduce the chances of data inflation via multiple coding of fixations that may lead to Type I errors (Barr 2008), we coded the fixation data as 1 or 0 per window according to the presence or absence of a fixation on the target AOI. In the Color experiment, the prominent accent L + H* (as compared to H*) reliably increased the likelihood to fixate on the target in the 0-to-300 ms window, and showed a marginally reliable effect for 300-to-600 ms window (Table 2, left). Although the estimated coefficient for the L + H* trials was still positive, the effect was not reliable for 600-to-900 ms window. In the Size experiment, L + H* exhibited a robust effect in increasing the likelihood to fixate the target in 0-to-300, 300-to-600, and 600-to-900 ms windows (Table 2, right). Given that it generally takes approximately 200 ms to plan and execute eye movements to much simpler displays on the basis of speech input (Allopenna et al. 1998; Dahan et al. 2001a, 2001b), the early diversions between the L + H* and H* functions in the present data could not have resulted solely from a search triggered by the nouns' segmental information. Therefore, we maintain our previous view that the accentual prominence on the pre-nominal modifier is immediately processed to initiate

[1] The window size was determined based on the distribution of fixation duration observed across the two experiments. The average fixation duration was 185 ms ($SD = 123$) for the Color and 192 ms ($SD = 133$) for the Size experiment. Across the two experiments, 90% of the fixations lasted less than 300 ms.

[2] Due to the space limit, these mean function figures are not included here. These figures are available upon request to the first author.

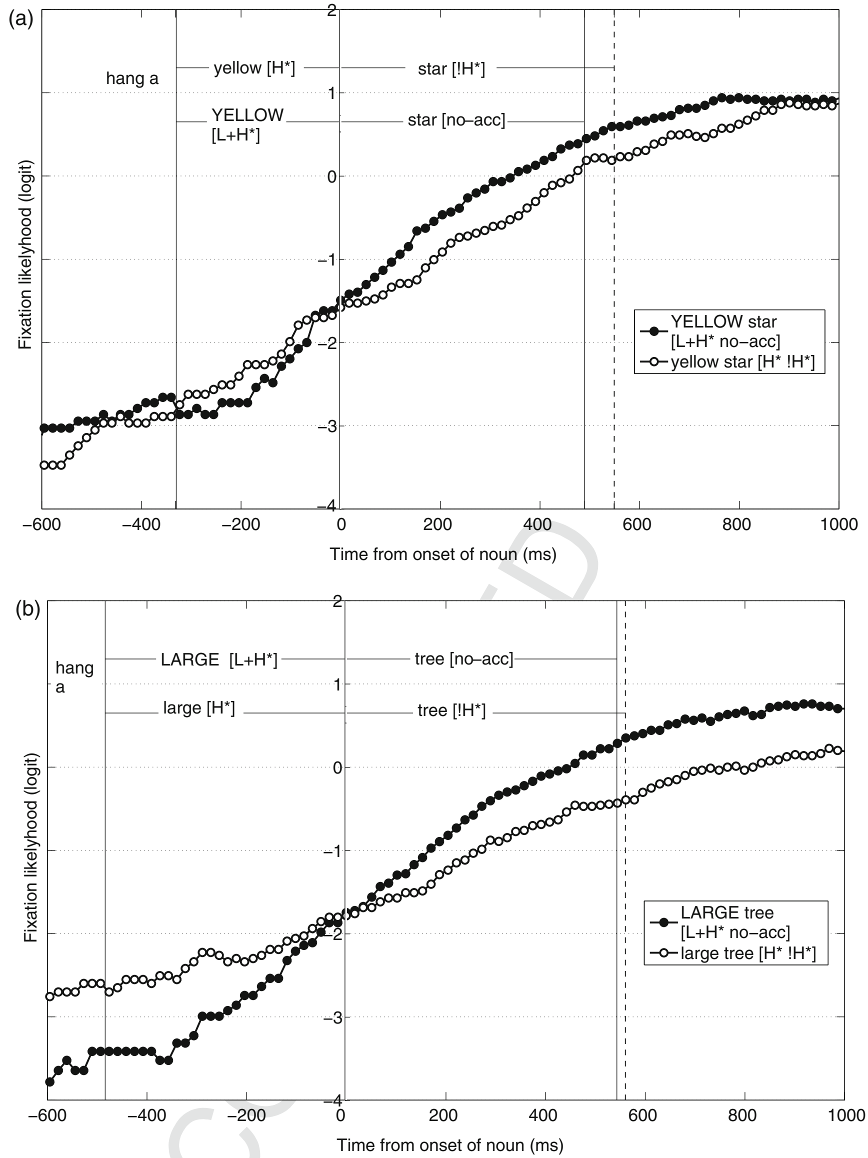


Fig. 3 Fixation likelihood (logit) function for the target ornament cell in the locally contrastive sequences in the (3a) Color and (3b) Size experiments

a search for the contrastive referent that is linked to the most salient or available entity in the listener's cognitive discourse structure. Such a claim by no means suggests that the following nouns' prosody had no contribution to comprehension of the discourse structure and to the visual search. Recall that the nouns tended to be shorter and had lower pitch after L + H* than after H* adjectives

Table 2 Experiment color & size: Effect of emphatic accent L + H* on the likelihood to fixate on the target

	Color ($N = 536$)				Size ($N = 637$)			
Time window from noun onset	Est.	*SE*	Wald Z	*p*	Est.	*SE*	Wald Z	*p*
−300 to 0	−0.019	(0.234)	−0.082	.935	−0.2	(0.224)	−0.89	.371
	(log-likelihood = −242.6)				(log-likelihood = −269.2)			
0 to 300	0.474	(0.179)	2.65	.008**	0.559	(0.175)	3.19	.0014**
	(log-likelihood = −357.8)				(log-likelihood = −390.7)			
300 to 600	0.336	(0.188)	1.78	.074^	0.677	(0.176)	3.86	.0001***
	(log-likelihood = −332.9)				(log-likelihood = −390.7)			
600 to 900	0.16	(0.203)	0.79	.431	0.57	(0.181)	3.14	.0017**
	(log-likelihood = −295.3)				(log-likelihood = −369.3)			

***$p < .001$, **$p < .01$, *$p < .05$, ^$p < .1$

in both experiments (Table 1). Thus, the faster increase in the fixations to the target may have been accelerated by the deaccentuation or prosodic attenuation of the following nouns that unfolded along with their segmental information.

The present results eliminate the possibility that our previous findings on the facilitative effect of L + H* were mere artifacts of the visual displays that benefited the eye movements returning to the cell fixated upon by the immediately preceding trial. However, the overall timing of the fixation rises in the present data were indeed slower than those in Ito and Speer (2008), where the fixation proportions reached their ceiling level (70–75%) by the end of the noun (whose average duration was 413 ms). In the present data, the fixations kept increasing throughout the nouns (average duration: 490 ms) in both experiments, and their ceiling levels were reached a few hundred milliseconds after the offset of the nouns. Since the overall number of ornaments and cells on the board were grossly comparable between the previous and the present studies, we attribute the differences in fixation timings to the more laborious search for the ornaments across the novel locations in the present study. Given the requirement for such additional effort in visual search, the consistent early emergence of the prosodic effect on the fixation rise should increase our confidence in listeners' ability to use intonation to anticipate upcoming referents.

The replication of the facilitative effect of L + H* in both the Color and the Size experiments also eliminates the possibility that the prosodic manifestation of contrast is redundant for subsective adjectives. In fact, the present data indicates no inherent semantic advantage for the subsective adjectives. If the use of size adjectives had automatically evoked the notion of contrast *without* the help of L + H* accent, the fixations to the target in the H* trials should have shown a faster increase with the size adjectives than with the color adjectives. However, the comparison of the slopes between Fig. 3a and 3b suggests the opposite – that is, the fixation increase in H* trials was relatively faster with the colors than with the sizes. Since prosody for the instructions in H* trials was controlled to not highlight any particular part of speech, these trials served as the baseline for the comparison within each sequence type. In other words, the fixation patterns from the H* trials reflect the general level of difficulty of visual search in each display. The data suggest that finding a cell on the board that contained objects in the same color but different sizes was generally more challenging than finding objects on the board with the same size but in different colors.

3.2 L + H Garden-Paths Visual Search?*

In the locally non-contrastive sequences, the presence of L + H* on the pre-nominal modifier led to numerically higher fixation likelihoods for the contrastive competitors as compared to H* trials in both the Color and the Size experiments. Again, the semantic advantage for the subsective size adjectives was absent from the H* trials. Contrary to our prediction, the likelihood to fixate on the contrastive competitor was generally much higher for the Color than for the Size experiment. Fig. 4a and 4b respectively show the likelihood to fixate on the target and the contrastive competitor in the L + H* and H* trials in the locally non-contrastive sequences in the Color experiment. When the pre-nominal adjective carried L + H* (Fig. 4a), the fixations to the contrastive competitor (e.g., 'green tree' in *red tree* → *GREEN ball*) competed with the fixations to the target until past 200 ms into the noun, despite the presence of the incoming noun's segmental information that could have been used to single out the target. To our surprise, there was also an increase in the looks to the competitor until beyond 200 ms into the noun in the H* trials (Fig. 4b). In both conditions, the rise in the fixations to the contrastive competitor was initiated before the critical noun. Such early increases in the fixations to the contrastive competitor indicate that participants in the Color experiment often fixated on the ornaments that stood in contrast with the preceding ornaments, regardless of the accentual pattern of the adjective. The timing of the declines in the looks to the contrastive competitor in Fig. 4a and 4b confirm that it took approximately 200 ms to re-direct the saccades on the basis of segmental information about the noun.

Due to this unexpected bias toward fixation on the contrastive ornaments, the accentual prominence on the pre-nominal color adjective led to a numerically larger but statistically non-significant increase in the likelihood to fixate on the contrastive competitor as compared to the non-prominent trials. Fig. 4c directly compares the fixations to the contrastive competitor between the L + H* and the H* trials within the Color experiment. Although L + H* resulted in a

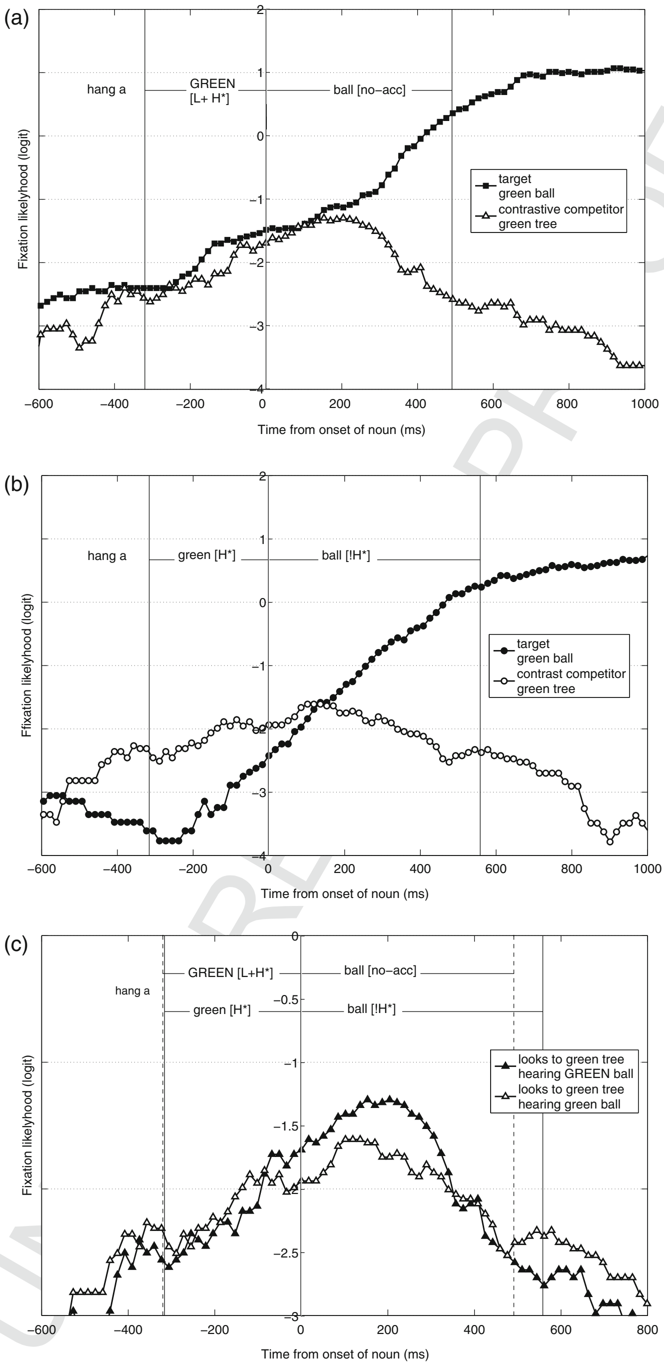


Fig. 4 Fixation likelihood (logit) functions for the target and for the contrastive competitor in the locally non-contrastive sequences in the (4a) L + H* trials and (4b) H* trials in the Color experiment. The fixation likelihood for the contrastive competitor directly compared between the L + H* and H* trials (4c)

higher likelihood of fixations to the contrastive competitor than H* in sequences such as *red tree* → *GREEN/green ball*, the mixed logistic regressions with a 300 ms window from -300 till 600 ms did not show any reliable differences[3] (Table 3, left). As participants were already frequently attending to the locally contrastive ornaments in separate locations in the Color experiment, the presence of accentual prominence on the pre-nominal adjective did not drastically enhance the interpretation of contrast, leading to only slightly more frequent fixations to the competitor.

The effect of accentual prominence on the pre-nominal adjective was far more visible in the Size experiment. Figures 5a and 5b respectively compare the fixations to the contrastive competitor (e.g., 'large tree' in *medium tree* → *LARGE/large ball*) and to the target (e.g. 'large ball') in the L + H* and the H* trials. As in the locally-contrastive sequences, the overall timing of the fixation increase was much slower in the Size experiment than in the Color experiment. Nonetheless, when the pre-nominal adjective had L + H*, there was an increase in the fixations to the contrastive competitor that lasted up to halfway into the noun, as shown in Fig. 5a. Note that the rise was initiated slowly, but started appearing within the adjective. In contrast, there was almost no increase in the fixations to the contrastive competitor when the adjective had H* (Fig. 5b). This indicates that the use of size adjectives did not automatically evoke a scalar-based contrast interpretation that would have directed the eyes toward ornaments that contrasted with the preceding one. The direct comparison of the fixations to the contrastive competitor between the two conditions confirmed the garden-path effect of L + H* (Fig. 5c). The mixed logistic regressions indicated reliably higher estimates of the likelihood to fixate on the contrastive competitor for L + H* as compared to H* for all three windows (Table 3, right).

Table 3 Experiment color & size: Effect of emphatic accent L + H* on the likelihood to fixate on the **Contrastive competitor**

	Color (N = 540)				Size (N = 638)			
Time window from noun onset	Est.	*SE*	Wald Z	*p*	Est.	*SE*	Wald Z	*p*
−300 to 0	−0.267	(0.203)	1.314	.189	0.89	(0.237)	3.75	.0001***
	(log-likelihood = −295.9)				(log-likelihood = −260.1)			
0 to 300	0.307	(0.196)	1.566	.117	0.953	(0.236)	4.04	5.25e-05***
	(log-likelihood = −312.4)				(log-likelihood = −263.7)			
300 to 600	0.144	(0.217)	.662	.508	0.596	(0.221)	2.7	.0068***
	(log-likelihood = −270.4)				(log-likelihood = −280.4)			

***p< .001, **p<.01, *p<.05, ^p<.1

[3]The mean likelihood functions for the contrastive competitors showed the increase and the decrease within -300-to-600 ms window in both experiments.

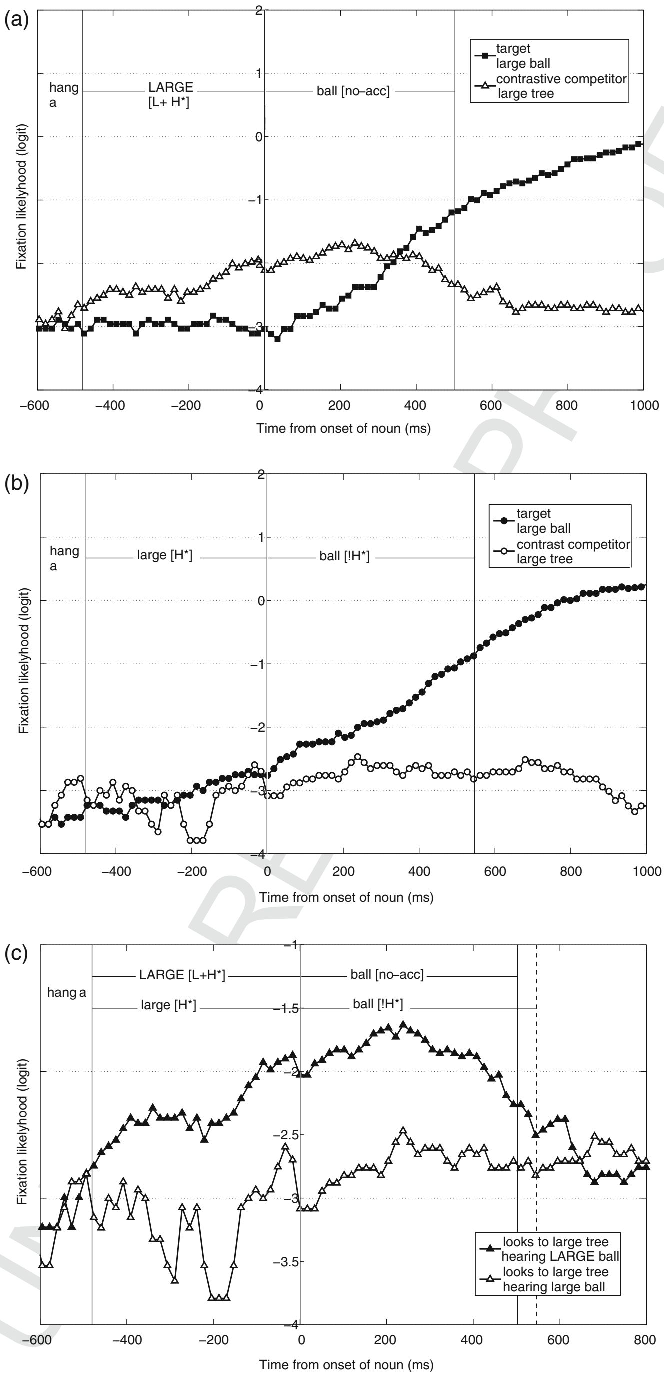


Fig. 5 Fixation likelihood (logit) functions for the target and for the contrastive competitor in the locally non-contrastive sequences in the (5a) L + H* trials and (5b) H* trials in the Size experiment. The fixation likelihood for the contrastive competitor directly compared between the L + H* and H* trials (5c)

3.3 Felicitous vs. Infelicitous Use of L + H*

Although the garden-path effect of L + H* was not shown as robustly as in our previous study, the present data confirmed the clear effect of discourse context on the interpretation of salient pitch accent. That is, the same prosodic contour [L + H* no-accent] yielded very distinct eye movements depending on the discourse context in which it appeared. In both experiments, the fixations to the target in the non-contrastive L + H* sequences were delayed as compared to those in the locally contrastive L + H* sequences. In the Color experiment, the fixations to the target in the non-contrastive sequences (Fig. 4a) did not rise as sharply from the noun onset as in the contrastive sequences (Fig. 3a, filled circle). Likewise, the fixation rise for the target in the Size experiment was evidently slower in the non-contrastive L + H* sequences (Fig. 5a) than in the contrastive L + H* sequences (Fig. 3b, filled circle). The mixed logistic regressions revealed reliably lower estimates of the likelihood to fixate on the target when L + H* was used infelicitously in non-contrastive sequences than when it was used felicitously in contrastive sequences for 0-300 ms in the Color, and for all four time windows in the Size experiment (Table 4).

In sum, the present data demonstrate that accentual prominence is immediately evaluated against the discourse context; when it is used felicitously, L + H* on the adjective and attenuation to the following repeated noun facilitated the detection of the target ornament. In contrast, the infelicitous use of L + H* before a noun that differed from the previous trial produced a false-alarm contrast, hampering the detection of the correct target.

Table 4 Experiment color & size: Effect of infelicitous context (Non-contrastive L + H*) on the likelihood to fixate on the **target**

	Color (N = 537)				Size (N = 635)			
Time window from noun onset	Est.	*SE*	Wald Z	*p*	Est.	*SE*	Wald Z	*p*
−300 to 0	0.14	(0.111)	1.263	.207	−0.319	(0.134)	−2.377	.0174**
	(log-likelihood = −262.2)				(log-likelihood = −209.3)			
0 to 300	−0.254	(0.089)	−2.847	.0044**	−0.988	(0.112)	−8.824	<2e−16***
	(log-likelihood = −359.2)				(log-likelihood = −312)			
300 to 600	−0.029	(0.096)	−0.301	.763	−0.768	(0.093)	−8.286	<2e−16***
	(log-likelihood = −323)				(log-likelihood = −373.1)			
600 to 900	−0.003	(0.104)	−0.027	.979	−0.558	(0.09)	−6.191	5.99e−10***
	(log-likelihood = −287.4)				(log-likelihood = −378.4)			

***p < .001, **p < .01, *p < .05, ^p < .1

4 Discussion

In the present study, the two different arrangements of identical sets of real-world objects uncovered a striking effect of visual and discourse context rather than of inherent adjective semantics on the use of prosodic prominence for referential resolution. Although the effects were weaker, the present findings generally supported our previous claim that accentual prominence on a pre-nominal adjective evokes an interpretation of contrast for the incoming referent, as the facilitative effect of L + H* in contrastive sequences and the numerical trend of the garden-path effect in the non-contrastive sequences were observed in both experiments. The fast increase in the fixations to the target in the contrastive sequences and the minor increase in the fixations to the contrastive competitor in the non-contrastive sequences may also have been enhanced by the prosodically attenuated nouns that followed the L + H*-accented adjectives, as the lack of accentual prominence on the noun may have led to the anaphoric interpretation of the referent (Dahan et al. 2002). Note, however, that extreme prosodic attenuation on the head noun is generally licensed by the presence of the preceding pitch prominence. A separate investigation may be required to test whether the interpretation of deaccentuation is independent from the degree of preceding accentual prominence, but we speculate that the post-L + H* attenuation conveys acoustic cues that are distinct from the simple production of a H*-accented word, and thus hypothesize that the prosodic status of the noun is incrementally evaluated against the prosodic status of the preceding adjectives. Follow-up research that examines the scope of prosodic prominence is underway.

The present data also echoed those of Ito and Speer (2008) in demonstrating that contrast-evoking accentual prominence had an opposite impact on referential resolutions, depending on the discourse context in which it appeared. When used felicitously, L + H* speeded the detection of the contrastive referent, while the infelicitous occurrence of L + H* resulted in a slower detection of the correct target. In other words, the prosody-driven garden-path effect surfaced more evidently as a delay in fixating on the correct non-contrastive targets than as an increase in the fixations to the incorrect contrastive referents. Recall that both effects were robust in our previous study where the ornament layout equated the detection of the target in contrast with the previous target to re-fixating the previously visited cell. In the present visual layout that demanded the search for an ornament in a novel location in each trial, the false-alarm L + H* did not always direct the eyes straight to the contrastive competitor, yet it certainly hindered the execution of saccades to the correct referents. Since it was the search strategy-related complexity (i.e., where contrastive things are located) and not the surface complexity (i.e., numbers of cells and ornaments) of visual environment that modulated the garden-path effect, we propose that the processing of accentual prominence is guided not only by the discourse context, but also by the task-relevant referential context of the visual field.

The idea that the interpretation of the accentual prominence depends on the referential context is directly supported by the differences in fixation patterns between the Color and the Size experiments. Contrary to our hypotheses, neither the comparison for the facilitative effect nor the comparison for the garden-path effect provided evidence for an automatic contrastive interpretation of the subsective size adjectives. Such a semantically dependent interpretation was predicted to appear as a fast increase in the fixations to the contrastive target for the locally contrastive H* trials and in a visible increase in the fixations to the contrastive competitor for the locally non-contrastive H* trials in the Size experiment. Instead, the fixations to the target in the contrastive sequences rose more quickly in the Color than in the Size H* trials, and a prosody-independent increase in fixations to the contrastive competitor in the non-contrastive sequences was observed in the Color, and not in the Size experiment.

We believe that such unexpected fixation patterns resulted from the difference in the ease of visual resolution between the two types of displays. As shown in the generally slower fixation increases in the Size experiment, detection of target ornaments among the competitors that shared color but differed in size appeared fairly challenging. During the data collection, we observed more frequent saccades and fixations comparing the ornaments across the cells in the Size than in the Color experiment. (Since the fixation data were collected only from the beginning of critical verb till 1000 ms after the end of the critical noun for each trial, the eye movements outside these windows that would reflect such tendency could not be analyzed quantitatively.) Also, more mistakes were made in the selection of ornaments in the Size (12) than in the Color (6) experiment. Such a difference in the difficulty of visual search was not anticipated, given that the two experiments arranged the identical sets of objects to display equal number of ornaments on each board, and that each board stayed in front of the participant until the tree was completed with the 26 ornaments. This long exposure to the visual targets was expected to counteract the complexity of the displays, making the search only moderately demanding for both of the experiments. However, the differences in size seemed to require more effortful search throughout the experiment, while the vivid differences in hue seemed to ease the detection of target ornaments, so much so that the eyes were swiftly sent to the contrastive competitors upon listening to the prosodically non-prominent (i.e., H*-accented) color adjective.

Basic research on visual processing adds insights to our discussion on the relationship between the visual environment and speech processing. In her extensive work on visual memory, Treisman (1998, 2006; Wheeler and Treisman, 2002) proposes that feature binding (e.g., between color and shape) is not automatic but requires task-dependent attention, and retention of bound information becomes harder as the display becomes more complex. Our search task with adjective – noun instructions forced bindings of color and shape and of size and shape. While studies that directly compare visual memory capacity or efficiency of visual search across different types of feature combinations are rare, Olsson and Poom (2005) suggest that the size-ratio of similarly shaped objects is more coarsely represented than color distinctions with easily categorized verbal labels (e.g., red, green, yellow). Although additional evidence is needed to confirm poorer representation of relative size as compared to distinct color, the difference in the ease of representation may explain the overall slower detection of the objects in the Size experiment. It is possible that the prosodic cues aided the feature binding effort more for visual searches that required extra attention due to coarse representation. If so, a color-based search might also benefit more from prosodic cues if the colors are not primary, but instead atypical shades crossing the natural boundaries for color perception (which hampers change detection (Olsson and Poom 2005).

In the present study, the color-sorted displays seems to have evoked a stronger notion of contrast than the size-sorted displays, and this scene perception seems to have biased the interpretation of incoming speech. In other words, the comprehension of auditory instructions was guided by the preceding scene interpretation. The effect of scene perception on language processing has been discussed in Altmann and Kamide (2004), who argue that the viewers can compute the likelihood of multiple possible scenarios for the given scene, and that linguistic inputs are integrated for dynamic update of the scene interpretation. In our experiments, visually clearer displays were definitely advantageous for the interpretation of each instruction. Although unintended, the present results illustrate that the experimental paradigm needs to be designed with care, as the eye movements are very sensitive to the referential context within the visual field.

Since the Size displays did not induce the predicted contrast bias, the two hypotheses that distinguish the additive and the complementary effect of prosody could not be attested as we planed. Nonetheless, the facilitative effect of L + H* in the contrastive sequences and the numerically higher fixations to the contrastive competitors in the non-contrastive sequences in the Color experiment demonstrated that prosody impacts the interpretation of utterances even when the referential relations are highly biased toward contrast. At least, it is not the case that accentual prominence is used complementarily only when no other cues to referential contrast are available. However, it remains unclear what exact perceptual factors are responsible for the interpretation of contrast. In particular, further analyses are necessary to determine what phonetic aspects of the prosodic prominence lead to the perception of contrastiveness. In order to set off this additional investigation, we conducted a series of step-wise multiple regressions, where four phonetic measures listed below and three random factors (subjects, color and size) were included as the predictors of the latency of the first correct fixations to the target among the felicitous L + H* trials. The first fixation latency was assumed to inversely reflect the ease of the target detection. The four phonetic measures entered in the analyses were: the F0 peak height of the adjective, the latency of the adjective F0 peak measured from the onset of the stressed syllable, the adjective duration, and the difference in the adjective F0 peak height between the target and the preceded utterances. These measures were selected to examine whether the contrastive interpretation of L + H* was triggered primarily by the acoustic variation within the target utterances or whether it was also prompted by the relative prominence of the target adjective gauged against the preceding utterance.

The results of these supplemental multiple regressions suggest that listeners may assess the prosodic prominence of the incoming speech with respect to the previous utterances. Among the four phonetic measures, the only phonetic factor that remained as a reliable predictor of the first fixation latency across the two experiments was the difference in the adjective F0 peak between the target and the previous utterances (Color: $t = -2.91$, $p<.01$, Total $R^2 = .051$, $F(2, 256) = 7.96$, $p<.001$; Size: $t = -5.66$, $p<.001$, Total $R^2 = .245$, $F(2, 296) = 25.38$, $p<.0001$. Note that the actual F0 peak height for the adjective also remained as a reliable predictor in the Size experiment: $t = 4.44$, $p<.001$). Within these initial analyses, the other factors such as color, size, and the subject did not appear as consistent predictors.

Since the prosodic patterns of our auditory stimuli were strictly controlled through the ToBI annotations and multiple perceptual screenings, the acoustic variations entered in the multiple regression analyses were relatively small, perhaps too much so to predict fine-grained differences in first fixation latency. Given the narrow range of prosodic manifestation, it was striking that the F0 difference between the adjacent instructions was the consistent predictor of the speed of target detection. Additional work is certainly required to test whether such relative prominence across utterances may guide the interpretation of accentual prominence when the instructions exhibit a wider variation in duration and F0 height, such as that found in natural speech. Also, future analyses might examine the contribution of other acoustic factors such as pitch excursion size (of both rise and fall, scaled in both F0 and semitones), intensity of the stressed syllable, vowel quality of the accented syllable, etc. To better quantify the values of these various phonetic cues, a perception study using gated stimuli may be useful (e.g., Petrone and D'Imperio, in this volume). In such an effort, a variety of dependent measures might also be explored. For example, the slope of the fixation rise may be more closely linked to the detection of the target than the first fixation latency, which may include the accidental fixations to the target cells. Therefore, future studies demand novel quantifications of eye movement data as well as finer-grained acoustic analyses of less controlled stimuli.

Finally, it needs to be clarified that the present experimental paradigm was not developed to demonstrate the categorical distinction between the two accent types – L + H* and H*. While the wide use of ToBI annotations has spread the assumption about the distinction between the two accents, a plausible phonetic and phonological continuum between H* and L + H* has been argued and experimentally tested by previous work such as Bartels and Kingston (1994), Ladd and Morton (1997), and Ladd and Schepman (2003) and Watson et al. (2008). An application of ToBI to annotate corpus data by Brugos et al. (2008) suggests that even trained annotators often experience the ambiguity between H* or L + H*, making the 'Alternative' tone tier useful. In Ladd and Morton (1997), participants assigned the degree of emphasis in a gradual manner to the stimuli with multiple ranges of F0 peak, while the results from the forced-choice interpretation and the classic same/different discrimination tasks suggested a categorical boundary in the mid F0 range that divides the pitch prominence into the emphatic and non-emphatic groups. These results led the authors to conclude that the pitch ranges are 'categorically interpreted, though they may not be categorically perceived' (Ladd and Morton 1997: 339).

Since the present study aimed to confirm the anticipatory contrast-evoking effect of distinctively prominent accent, which is conventionally labeled as L + H* as opposed to H*, we used the canonical renditions of the accents that represented the two ends of the prominence continuum. Thus, the effect of accentual prominence was sought by comparing trials with two extreme F0 contours (flat line vs. evident excursion). Obviously, finding the differences in the fixation patterns between these conditions does not clarify whether the accentual prominence is distributed in a categorical manner in natural productions, let alone whether intermediate prominence triggers categorical or gradient responses. In spontaneous utterances, intermediate F0 levels are expected to be ubiquitous, and we suspect that they are interpreted according to the referential clarity within the scene and the discourse status of the accented entities. It is a very interesting empirical question whether the eye movements remain as sensitive measure of such a context-dependent processing of gradient accentual prominence.

5 Conclusion

Using eye movements as a continuous online measure of responses to incoming speech, the present study demonstrated that prosodic prominence is interpreted according to referential and discourse contexts, rather than the inherent semantic properties or the absolute acoustic properties of particular words. While a prominent accent seemed to evoke contrast in a non-complementary manner, comparable pitch prominences may have opposite effects on the interpretation of utterances, depending on the informational status of references that must be constantly updated as a discourse proceeds. As the exact acoustic features that trigger the interpretation of contrast remain unknown and a categorical distinction among accents with seemingly-gradual ranges of pitch height and alignment has yet to be verified, we hope to continue developing our eye-tracking paradigm and the analytical tools to investigate responses to prosodic prominence in more naturalistic utterances in the future.

Acknowledgements The present research was supported by NIH grant DC007090. We thank Ping Bai for assistance with data analysis, and Laurie Maynell, Ross Metusalem, and Julie McGory for assistance with creating and ToBI-annotating our spoken stimuli, and the anonymous reviewers for helpful comments and suggestions.

References

Allopenna, Paul D., James, S. Magnuson, and Michael K. Tanenhaus. 1988. Tracking the time course of spoken word recognition using eye movements: evidence for continuous mapping models. *Journal of Memory and Language* 38: 419–439.

Altmann, Gerry T. M. and Yuki Kamide. 2004. Now you see it, now you don't: Mediating the Mapping between Language and the Visual World. In John M. Henderson and Fernanda Ferreira (eds.) *The Interface of Language, Vision, and Action*, 347–383. New York: Psychology Press.

Barr, Dale J. 2008. Analyzing 'visual world' eyetracking data using multilevel logistic regression. *Journal of Memory and Language* 59: 457–474.

Barr, Dale J. and Austin F. Frank. 2009. Analyzing multinomial and time-series data. *Workshop on Ordinary and Multilevel Modeling at 2009 CUNY Conference on Sentence Processing, UC Davis*. March, 25, 2009.

Bartels, Christine and John Kingston. 1994. Salient pitch cues in the perception of contrastive focus, In Paul Bosch and Rob van der Sandt (eds.) *Focus and natural language processing. IBM Working Papers on Logic and Linguistics, v. 6*, 1–10. Heidelberg.

Beckman, Mary E. and Gayle M. Ayers. 1997. *Guidelines for ToBI labelling, vers 3.0* [manuscript]: Ohio State University.

Beckman, Mary E., Julia Hirschberg and Stefanie Shattuck-Hufnagel. 2005. The Original ToBI System and the Evolution of the ToBI Framework. In Sun-Ah Jun (ed.) *Prosodic Typology: The Phonology of Intonation and Phrasing*, 9–54. New York: Oxford University Press Inc.

Bolinger, Dwight L. 1961. Contrastive accent and contrastive stress. *Language* 37: 83–96.

Brugos, Alejna, Nanette Veilleux, Mara Breen and Stefanie Shattuck-Hufnagel. 2008. The Alternatives (Alt) Tier for ToBI: Advantages of Capturing Prosodic Ambiguity. Poster presented at *4th Conference on Speech Prosody 2008*, May, Brazil. http://aune.lpl.univ-aix.fr/~sprosig/sp2008/papers/id072.pdf

Chierchia, Gennaro and Sally McConnell-Ginet. 2000. *Meaning and Grammar: An Introduction to Semantics*. Cambridge: MIT Press.

Chierchia, Gennaro and Raymond Turner. 1988. Semantics and Property Theory. *Linguistics and Philosophy* 11: 261–302.

Dahan, Delphine, James S. Magnuson, and Michael K. Tanenhaus. 2001a. Time course frequency effects in spoken-word recognition: evidence from eye-movements. *Cognitive Psychology* 42: 317–367.

Dahan, Delphine, James S. Magnuson, Michael K. Tanenhaus and Ellen M. Hogan. 2001b. Subcategorical mismatches and the time course of lexical access: evidence for lexical competition. *Language and Cognitive Processes* 16: 507–534.

Dahan, Delphine, Michael K. Tanenhaus and Craig G. Chambers. 2002. Accent and reference resolution in spoken-language comprehension. *Journal of Memory and Language* 47: 292–314.

Henderson, John M. and Fernanda Ferreira. 2004. Scene Perception for Psycholinguists. In John M. Henderson and Fernanda Ferreira (eds.) *The Interface of Language, Vision, and Action*, 1–58. New York: Psychology Press.

Ito, Kiwako, Nobuyuki Jincho, Utako Minai, Naoto Yamane, and Reiko Mazuka. (ms. under revision). Intonation facilitates contrast resolution: Evidence from Japanese adults & 6-year olds.

Ito, Kiwako and Shari R. Speer. 2008. Anticipatory effect of intonation: Eye movements during instructed visual search. *Journal of Memory and Language* 58: 541–573.

Ito, Kiwako, Nobuyuki Jincho, Naoto Yamane, Utako Minai and Reiko Mazuka. 2009. Use of emphatic pitch prominence for contrast resolution: An eye-tracking study with 6-year old and adult Japanese listeners. Paper presented at *Boston University Conference on Language Development 34*, Boston, MA.

Jaeger, T. Florian. 2008. Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language* 59: 434–446.

Johnson, Keith. 2008. *Quantitative Methods in Linguistics*. Oxford: Blackwell.

Kamp, Johan. A. W. 1975. Two Theories about Adjectives. In Edward Keenan (ed.) *Formal Semantics of Natural Language*, 123 – 155. Cambridge: Cambridge University Press.

Kamp, Hans and Barbara Partee. 1995. Prototype theory and compositionality. *Cognition* 57: 129–191.

Ladd, D. Robert and Rachel Morton. 1997. The perception of intonational emphasis: continuous or categorical? *Journal of Phonetics* 25: 313–342.

Ladd, D. Robert and Astrid Schepman. 2003. Sagging transitions' between high pitch accents in English: experimental evidence. *Journal of Phonetics* 31: 81–112.

Metusalem, Ross and Kiwako Ito. 2008. The role of L + H* pitch accent in discourse construction. *Proceedings of Speech Prosody 2008, an international conference*, Campinas, Brazil. May. http://aune.lpl.univ-aix.fr/~sprosig/sp2008/papers/id142.pdf

Olsson, Henrik. and Leo Poom,. 2005. Visual memory needs categories. *Proceedings of the National Academy of Science* 102 (24): 8776–8780.

Oltean, Stefan. 2007. On the Semantics of Adjectives. *Studia Universitatis Babes-Bolyai – Phlologia*: 155–164.

Pierrehumbert, Janet and Julia Hirschberg. 1990. The meaning of intonational contours in the interpretation of discourse. In Phillip Cohen, Jerry Morgan and Martha E. Pollack (eds.) *Intentions in Communication*, 342–365. Cambridge: MIT Press.

Sedivy, Julie, Michael K. Tanenhaus, Craig G. Chambers, and Greg Carlson. 1999. Achieving incremental semantic interpretation through contextual representation *Cognition* 71: 109–147.

Treisman, Anne. 1998. Feature binding, attention and object perception. *Philosophical Transactions of the Royal Society, Series B* 353: 1295–1306.

Treisman, Anne. 2006. Object tokens, binding and visual memory. In *Handbook of Binding and Memory: Perspectives from Cognitive Neuroscience*, eds. Hubert Zimmer, Axel Mecklinger, and U.lman Lindenberger, 315–338. New York, Oxford University Press.

Watson, Duane G., Michael K. Tanenhaus and Christine A. Gunlogson. 2008. Interpreting pitch accents in on-line comprehension: H* vs. L + H*. *Cognitive Science: A Multidisciplinary Journal* 32, 7: 1232–1244.

Weber, Andrea, Bettina Braun, Mathew W. Crocker. 2006. Finding Referents in Time: Eye-Tracking Evidence for the Role of Contrastive Accents. *Language and Speech* 49 (3): 367–392.

Wheeler, Mary E. and Anne M. Treisman 2002. Binding in Short-Term Visual Memory. *Journal of Experimental Psychology: General* 131 (1): 48–64.

The Developmental Path to Phonological Focus-Marking in Dutch

Aoju Chen

1 Introduction

In many languages focus is typically accompanied by certain intonational events. For this reason, focus is sometimes treated as an intonational category and the term 'focus' has been used to refer to accentuation, sentence stress or acoustic prominence in the literature (Féry 2007). In this paper, focus is considered an information structural category (or a primitive of information packaging) only and defined as the constituent that carries the new information in a sentence to the addressee (e.g. Lambrecht 1994; Gundel 1999). Focus is often discussed in theories of information structure together with the concept 'topic'. Topic refers to the discourse entity about which the new information is provided. Focus becomes contrastive if the new information conveyed is chosen from a closed set of alternatives in the discourse (Chafe 1974). It can also have different scopes, i.e. a single lexical word (narrow focus) vs. more than one lexical word (broad focus) (Ladd 1980). Contrastive focus usually has a narrow scope. In West Germanic languages and some Romance languages, both the placement of pitch accent and the type of pitch accent (i.e. the phonological cues) are essential to the marking of focus. Further, gradient variations in pitch, duration and peak alignment (i.e. the phonetic cues) also play a role, in particular in distinguishing different focus conditions, i.e. broad focus, narrow focus, contrastive focus, which tend to be marked with similar phonological cues (e.g. Baumann et al. 2007; Hanssen et al. 2008).

Research on early intonational development has shown that children acquiring West Germanic and Romance languages have developed the inventory of pitch accents and boundary tones in the adult model by the late two-word stage

A. Chen (✉)
Max Planck Institute for Psycholinguistics, P.O. Box 310, 6500, AH Nijmegen, The Netherlands
e-mail: aoju.chen@mpi.nl

S. Frota et al. (eds.), *Prosodic Categories: Production, Perception and Comprehension*, Studies in Natural Language and Linguistic Theory 82,
DOI 10.1007/978-94-007-0137-3_5, © Springer Science+Business Media B.V. 2011

(e.g. Balog and Snow 2007 for English; Chen and Fikkert 2007a for Dutch; Prieto and Vanrell 2007 for Catalan; Frota and Vigário 2008 for European Portuguese). The questions that arise are thus whether children can use phonological cues to mark focus from early on and if not, how they acquire phonological focus-marking over time. In this paper, to address these questions I will consider in detail three very recent studies on the phonological marking of non-contrastive narrow focus in Dutch children (study 1 reported in Chen and Fikkert 2007a, studies 2 and 3 reported in Chen in press). When necessary, I will present results from new analyses conducted for the purpose of the current paper. In the light of both reported and new results, I will show that learning to use phonological cues to mark focus is a gradual process, and put forward a first proposal on the developmental path to adult-like phonological focus-marking in Dutch. The three studies to be discussed are concerned with focus marking in both spontaneous and elicited production by typically developing monolingual Dutch children aged between 1;9 and 8;10.

In the rest of the introduction section, I first briefly consider phonological focus-marking in adult Dutch (Section 1.1) and then review past work on focus marking in child language (Section 1.2).

1.1 Phonological Focus-Marking in Adult Dutch

Chen (2007) examined the phonological marking of focus and topic in naturally spoken SVO sentences elicited as answers to WH-questions in adult Dutch. In half of the answer sentences, the subject NP was focused and the object NP was unfocused (i.e. the topic). In the other half of the answer sentences, the object NP was focused and the subject NP was unfocused. As regards the phonological marking of focus, she found that the noun of the focused NP was nearly always accented independent of the position of the NP in the sentence and the preferred accent type was H*L (a fall). In sentence-final position, the noun of the focused NP could also be realised with !H*L (a downstepped fall – a falling accent with a lowered pitch peak relative to the peak of the preceding accent). The noun of the unfocused NP was however, realised differently depending on the position of the unfocused NP relative to the focused NP. The pre-focus noun could be unaccented but was mostly accented, most frequently with H*L, like the focused noun. Following Horne (1990), Chen suggested that the placement of accent in the pre-focus noun was rhythmically motivated. That is, accenting the pre-focus noun led to the preferred strong-weak (in the verb)-strong rhythmic pattern. In contrast, the post-focus noun was preferably realised without accentuation, although it was sometimes realised with !H*L. The use of !H*L in the sentence-final noun (either focused or unfocused) was interpreted as the speaker' means to express lack of interest in further discussion on the point that she or he made in the utterance, as suggested by Gussenhoven et al. (2003).

1.2 Focus Marking in Child Language

Past work on the intonational realisation of focus in child language is rather limited and mostly concerned with contrastive focus or contrast. In a picture-description task, Hornby and Hass (1970) asked English three- to four-year-olds to describe pairs of pictures that differed by one feature (subject, verb, or object). They found that children frequently used ‘contrastive stress’ (i.e. emphatic accentuation in the form of a fall with a rather wide pitch range) to pronounce the word carrying the contrastive information in their description of the second picture, although they used contrastive stress even more frequently in subject contrast than in verb- and object-contrast. Using a similar method, MacWhinney and Bates (1978) found that the use of contrastive stress was well established around the age of three but still increased between three and six in English children Wells et al. (2004) examined how English children used intonation to express contrast in adjective + noun phrases. They found that five-year-olds could use accentuation to mark contrast, although there was misplacement of accent in non-phrase final contrast. The ability to express contrast intonationally by age five has also been reported for German children. Müller et al. (2006) elicited SVO sentences with a contrast either in the subject or in the object from German four- to five-year-olds by means of a question-answer task. In this task, children repeated a puppet’s answer to a question about a series of comic strips; the puppet’s speech lacked intonation and rhythmical properties. Müller et al. found that children, like adults, uttered the words carrying the contrast with a higher mean pitch than the words carrying no contrast with the same syntactic function and in the same sentence position (albeit differing segmentally). However, these studies tell us little about the types of accent that children use to mark contrast. Moreover, they have limited implications for children’s ability to use intonation in marking non-contrastive focus.

Wieman (1976) did discuss the use of accentuation in marking non-contrastive focus. She observed in spontaneously produced two-word utterances by two-year-old English children that accent placement in the two-word stage was in the first place governed by the semantic relation between the two words in an utterance. For example, in Verb-Locative utterances (e.g. play museum), the accent was almost always assigned to the locative (e.g. museum). However, the default pattern broke down if the non-default accent-bearing word carried new information or was in focus. For example, a child accented ‘firetruck’ in ‘firetruck street’ when answering his mother’s question ‘What is in the street?’. Wieman interpreted this observation as an indication that two-year-olds could strategically assign accents to mark non-contrastive focus. However, this result was based on seven sentences in her data only and its generalisability is therefore highly questionable (cf. Wells and Local 1993). In a case-study on the prosodic and syntactic organisation of a German-acquiring child’s two-word utterances, Behrens and Gut (2005) analysed the intonation of the child’s two-word utterances produced between 2;0 and 2;3. They found that the child frequently uttered both words with accentuation in this period of time. There is thus no conclusive evidence that children can use accentuation to mark non-contrastive focus in the two-word stage. Besides, Wieman (1976) did not consider children's choice of accent type in focus-marking.

Different from the earlier studies reviewed above, the three studies to be discussed here have examined both accent placement and choice of accent type in the marking of non-contrastive narrow focus (hereafter focus) in detail.

2 General Methodological Issues

In the three studies, the sentences included for intonational analysis were processed in a similar way. Specifically, a textgrid was created for each sentence in Praat (Boersma 2001). On the word tier (interval tier) of each textgrid, landmarks were inserted to demarcate the boundaries of each word and the words were transcribed orthographically. Then each sentence was intonationally transcribed on the intonation tier (point tier) following the notation Transcription of Dutch Intonation (ToDI) (Gussenhoven et al. 2003; Gussenhoven 2005) by the author or the first author without access to the context in which the sentence was produced. Alternative labels were noted down on the alternative tier (interval tier) in case of doubts. In ToDI, five basic accent types are recognised, H* (typically a high-pitch stretch or a rise without a distinct low plateau in the stressed syllable), L* (a low-pitch stretch), H*L, L*H (a rise), and H*LH (a fall-rise). In addition, there are modified versions of these accents, for example, !H* and !H*L - the downstepped H* and H*L.

As ToDI is developed for the purpose of transcribing intonation contours in adults' speech, applying ToDI to children's speech may run the risk of shoehorning children's intonation contours in adults' intonational categories. To minimise such a risk, a phonetic description using a ToDI-like label (e.g. H*L^HL depicting two distinct falls in a compound noun) was given when the shape of an accent did not fit with the description of any of the pitch accent types in ToDI. This however turned out to be necessary only in a very small number of cases. Moreover, extra symbols were introduced to code observable variations in pitch scaling and peak alignment in H*L and !H*L in utterances produced in the late two-word stage. This was done to find out whether these kinds of variations played a role in focus marking.

In Study 1, a second ToDI transcriber checked all accent labels and gave alternative labels in case of disagreement. In Studies 2 and 3, a second ToDI transcriber transcribed a subset of the data and checked the labels in the rest of the data. In all the three studies, intra-transcriber disagreements were resolved by the two transcribers together. Measures of inter-transcriber agreement were reported for the data of the older children in Chen (in press).

Accent labels (including 'no accent') were automatically extracted from each sentence using a Praat script and were subsequently subjected to descriptive analyses in Study 1 and statistical analyses in Studies 2 and 3.

3 Study 1: Two-Year-Olds

Chen and Fikkert (2007a) examined the effect of information status (new vs. given) on accent placement in two-word utterances produced by three children aged between 1;9 and 2;1 after the vocabulary size of 160 unique recorded words was reached (defined as the late two-word stage by the authors). The utterances were selected from the longitudinal data of these three children available in the CLPF database (Fikkert 1994; Levelt 1994). It was found that both words were accented in most of the utterances regardless of information status.

For the purpose of the current paper, I reanalysed the distribution of accent patterns in Noun-Verb utterances from the perspective of focus marking. As the utterances were mostly children's responses to an adult interlocutor's questions or comments about a toy or an ongoing activity in the direct surroundings, the focus in 20 of the 31 Noun-Verb utterances could be reliably identified in the corresponding context. My reanalysis was confined to these 20 utterances. In five of these utterances, the noun was focused (e.g. *appel eten* 'apple eat' uttered as the answer to the question 'what is the horse eating?'). The focused noun was accented in all five cases; the unfocused verb was accented in four cases. In another eight utterances, both the noun and the verb were in focus (e.g. *tanden poetsen* 'teeth clean' uttered as the answer to the question 'what is the boy doing?'). The focused noun and the focused verb were accented in all eight utterances, whereas an adult speaker would typically only accent the focused noun in such cases.[1] Six of the remaining seven utterances were repetitions of what an adult said. The verb was accented in all six utterances; the noun was accented in five of these utterances and was devoiced in one of these utterances. In the last remaining utterance the verb was in focus (i.e. *poes huilen*, 'cat cry', uttered as the answer to the question 'what is the cat doing?'). The focused verb was accented and the unfocused noun was devoiced. Taken together, we see that accentuation was mostly placed independent of focus and both words were accented in all but two utterances. The most common tunes in these utterances were H* !H*L and H*L !H*L followed by a low boundary tone. The phonetic realisation of H*L and !H*L played no role in focus marking. Thus, unlike adults, the children in the two-word stage did not use accent placement to mark focus, contra Wieman's (1976) claim but in line with Behrens and Gut's (2005) finding.

However, this may not be the whole picture on phonological focus-marking in two-year-olds. Children of this young age are known to have an immature pitch-control system. They might therefore experience difficulty with keeping pitch low over the length of the second word after the falling accent in the first word or with lowering the pitch to the baseline of their pitch range in the second word after a high-pitch stretch or a rise in the first word. This was in fact evidenced by their use of almost complete devoicing to realise an unfocused word (e.g. 'poes'

[1] A male Dutch speaker was asked to answer a number of questions in two words in comparable contexts as those in the CLPF corpus. As expected, the male speaker accented only the noun in the answer to the question 'what is the boy doing?'.

in *poes huilen*), as illustrated in Fig. 1(a). In this light, the use of !H*L could be considered an alternative strategy in addition to devoicing when ‘no accent’ should be produced. I then reanalysed the distribution of acoustic realisations in the Noun-Verb utterances by grouping !H*L and devoicing together as the acoustically weak patterns, standing in contrast to the acoustically strong patterns (i.e. the non-downstepped accents). Interestingly, a different picture then emerged. That is, the focused word was always accented with a non-downstepped accent (H*L or H*), whereas the unfocused word was mostly spoken with an acoustically weak pattern, as illustrated in Fig. 1(b) and 1(c). Further, when both the noun and the verb were focused, the noun was realised with a non-downstepped accent but the verb was realised with !H*L. These results thus show that in the late two-word stage, children systematically used

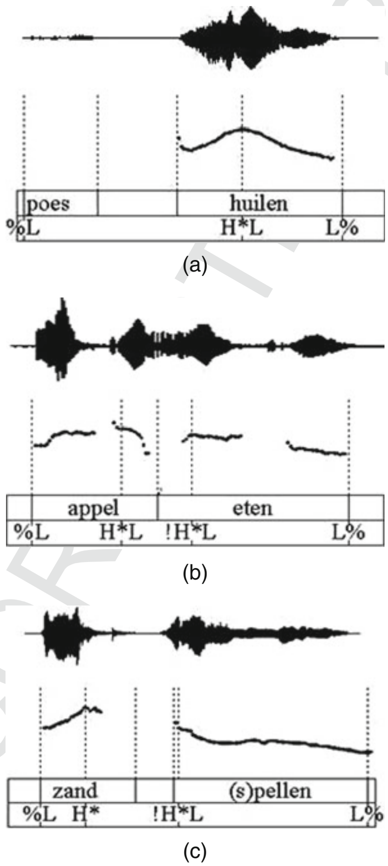


Fig. 1 Examples of H* and H*L observed in the focused words (*huilen* ‘cry’, *zand* ‘sand’, *appel* ‘apple’) and examples of devoicing and !H*L observed in the unfocused words (*poes* ‘cat’, *spellen* ‘play’, *eten* ‘eat’) in the two-word utterances produced by Dutch two-year-olds. The pitch contours were plotted in the range between 100 Hz and 550 Hz

non-downstepped accents to realise the focused word or the word that usually gets accented in adults' speech in the same focus condition, but used a downstepped accent or devoicing to realise the unfocused word and the word that usually does not get accented in adults' speech.

It is worth mentioning that the same patterns were also found in the Adjective-Noun utterances (e.g. *lieve beer* 'sweet bear') in Dutch three-year-olds. These utterances were produced as answers to questions about an attribute of the nouns (e.g. *Wat voor beer is hij?* 'What kind of bear is he?') (Chen and Fikkert 2007b).

However, it should be noted that the use of non-downstepped accents on the focused word in the two-word utterances may be confounded by the fact that downstepped accents usually do not occur on the first word and the focus was on the first word in some of the Noun-Verb utterances and all the Adjective-Noun utterances. Analysis on young children's two-word utterances with focus on the second word is thus called for to verify the observed relationship between focus and the acoustic strength of the production.

4 Study 2: Four- to Five-Year-Olds

Study 2 is concerned with four- to five-year-olds' phonological marking of focus and topic in full sentences (the 'neutral' four- to five-year-olds in Chen in press). In this section, I will discuss the relevant details of the study from the perspective of focus-marking.

SVO declaratives were elicited as answers to WH-questions about either the subject or the object. Both subjects and objects were realised with full NPs. In half of the SVO sentences, the sentence-initial NP (the subject) was focused and the sentence-final NP (the object) was unfocused (i.e. the topic). In the other half of the SVO sentences, the sentence-final NP was focused and the sentence-initial NP was unfocused. Each noun (e.g. *poetsvrouw* 'cleaning-lady') appeared in the focused NP in one answer sentence but in the unfocused NP in another answer sentence, as illustrated in (1).

(1) Experimenter: *Kijk! Een biet. Wie eet de biet?*
'Look! A beet. Who is eating the beet?'
Participant: *[De poetsvrouw]$_{focus}$ eet [de biet]$_{topic}$.*
'De cleaning-lady is eating the beet.'
Experimenter: *Kijk! Een poetsvrouw. Wat pakt de poetsvrouw?*
'Look! A cleaning-lady. What is the cleaning-lady picking up?'
Participant: *[De poetsvrouw]$_{topic}$ pakt [een vaas]$_{focus}$.*
'The cleaning-lady is picking up a vase.'

4.1 Data Elicitation

A picture-matching game was used to elicit the SVO sentences. Prior to the game, the experimenter showed each child two boxes full of pictures. The child was told that a picture from one box went together with a picture from the other box and that the experimenter needed his/her help to sort the pictures out. The procedure of the game is as follows. First, the experimenter took a picture (e.g. a picture of a cleaning-lady) from one box. She then drew the child's attention to the picture and established what the picture was by saying *Kijk! Een poetsvrouw!* 'Look! A cleaning-lady!' with either H*L or L*H on the verb and H*L on the noun. In the picture, the cleaning-lady seemed to be picking up something. The experimenter then asked a question about the picture (e.g. '*Wat pakt de poetsvrouw?* 'What is the cleaning-lady picking up?'), again in a prescribed intonation contour. The WH-word was spoken with H*L; the noun was spoken with either 'no accent' or !H*L. Second, the child turned to a robot for help by clicking on a picture of the robot displayed on his/her computer screen. The child received the answer (in SVO word order) from the robot via a headphone set such that the experimenter could not hear it.[2] Third, the child then used the same (lexical) words as the robot to answer the experimenter's question but in his/her own intonation (e.g. *De poetsvrouw pakt een vaas*. 'The cleaning-lady is picking up a vase.'). Finally, the experimenter looked for the matching picture from the other box and handed both pictures over to the child.

Twenty-eight four- to five-year-old monolingual Dutch children participated in the experiment. The children were tested individually in a quiet room at their school during school time. Each session was recorded with an external high-quality microphone connected to a portable DAT recorder at 48 kHz sampling rate with 16-bit resolution. The microphone was placed 10–15 cm away from the mouth of the children. Responses to 36 WH-questions were elicited from each child.

4.2 Intonational Analysis

A selection of the recordings was made on the basis of level of background noise, quality of segmental articulation, speaking style (neutral vs. playful), and whether the speaker had speaking or hearing deficits. In total, full-sentence responses from 12 children with a neutral speaking style (age range: 4;5 – 5;7, mean age 5;1) were intonationally transcribed. Some of the nouns in these sentences were not included for further analysis because of problems that could affect choice of intonation pattern, such as misplacement of word stress, false start, breaking a word into two parts, phrasing, and laughing while speaking. In total, the intonation patterns in 300 nouns in sentence-initial NPs

[2]The robot's answer sentence was generated by splicing together the words (with a 200 ms pause in between) recorded in a wordlist reading task. The original intonation was then erased and the pitch level was set at 200 Hz to obtain a flat intonation pattern.

(hereafter sentence-initial nouns) and 276 nouns in sentence-final NPs (hereafter sentence-final nouns) were included for further analysis.

4.3 Results and Discussion

Table 1 shows the mean percentage distribution of the intonation patterns in the nouns for each sentence position separately. The mean percentage distribution of a given intonation pattern in a given condition (focused vs. unfocused) was computed by averaging the percentages of the nouns spoken with that accent type in the respective condition from all the children. In sentence-initial position, both the focused noun and the unfocused noun were mostly accented, most frequently with H*L, followed by H*. Within the limited number of unaccented tokens, most of them occurred in the unfocused condition. In sentence-final position, the focused noun was mostly accented, most frequently with L*H, followed by !H*L and H*L (Fig. 2), whereas the unfocused noun was most frequently realised with ‘no accent’. But when the unfocused noun was accented, the most frequently used accent type was L*H, followed by !H*L and H*L. These observations suggested that information structure mattered to accent placement in both sentence positions but not choice of accent type.

To verify the observed relationship between intonation patterns and information structure statistically, Chen (in press) carried out multinomial logistic regression (hereafter MLR) modelling at the significance level of 0.05 on the intonation patterns in sentence-initial nouns and sentence-final nouns separately. In each model, the independent variable (or the predictor variable) was INFORMATION PACKAGING with two categories, topic (non-focus) and focus. The variable SPEAKER was used to define the subgroups of the data in the model. The dependent variable (or the outcome variable) was the intonation in the nouns, consisting of four categories (H*, H*L, OTHER, and ‘no accent’) in sentence-initial position and five categories (H*L, !H*L, L*H, OTHER, and ‘no accent’) in sentence-final position. The reference category was either ‘no accent’ or H*L.

Sentence-initial nouns: The MLR modelling showed that the model fitting was marginally significantly improved after the variable information packaging

Table 1 Mean percentage distributions of the intonation patterns in the nouns (four- to five-year-olds) (adapted from Table 2 in Chen in press)

		H*	H*L	!H*L	L*H	OTHER[a]	no accent
initial	focused	27.3%	55.6%	n.a[b]	n.a[b]	11.3%	5.8%
	unfocused	18.9%	61.8%	n.a[b]	n.a[b]	8.3%	11%
final	focused	n.a[c]	15.5%	33%	40.9%	2.1%	8.5%
	unfocused	n.a[c]	8.3%	28.5%	26%	3.6%	33.6%

[a]The category ‘OTHER’ refers to infrequently used accent types.
[b]The accent type did not occur in the condition to which the cell corresponds.
[c]The accent type was grouped into the OTHER category.

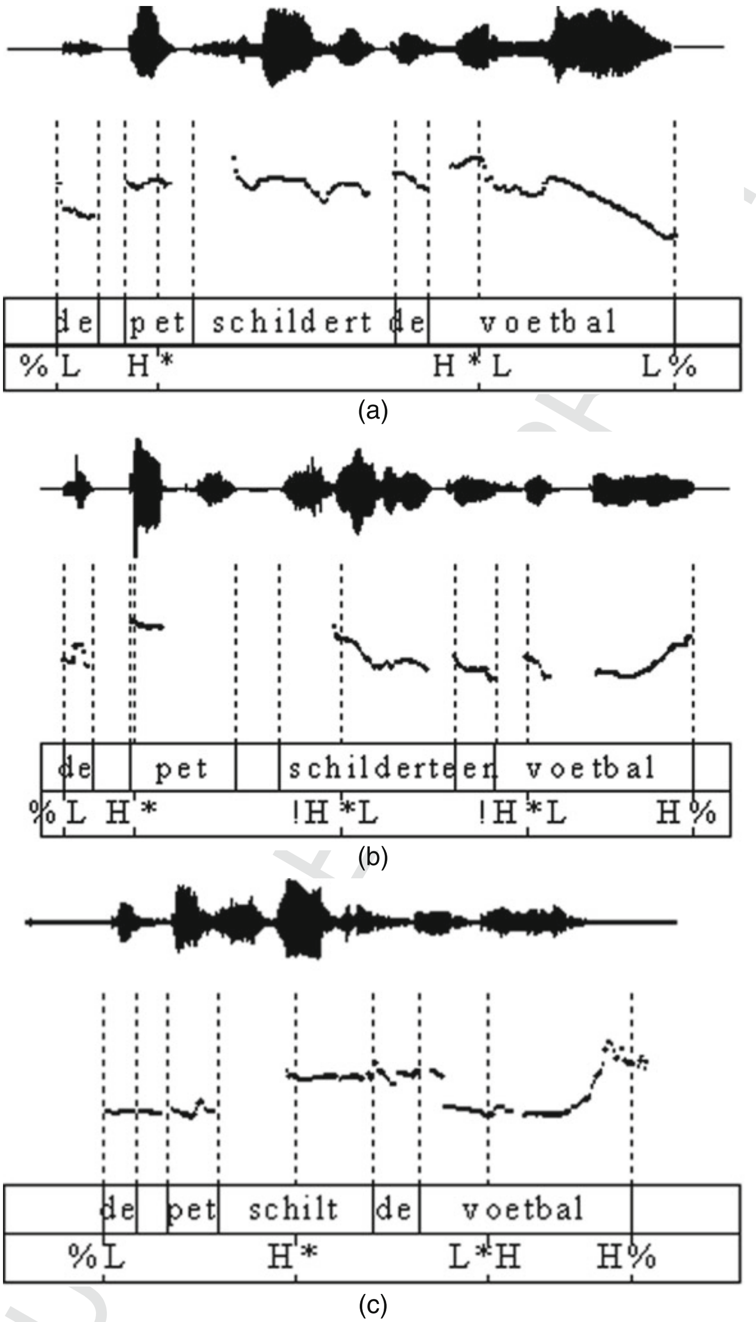


Fig. 2 Examples of H*L, !H*L, and L*H as realised in the word '*voetbal*' 'footbal' in the sentence '*De pet schildert een/de voetbal*' 'The hat is painting a/the football' produced by Dutch four- to five-year-olds with the sentence-final NP being focused. One child reduced 'schildert' to 'schilt'. The pitch contours were plotted in the range between 80 Hz and 450 Hz (Reproduced from Fig. 2 in Chen in press)

was added to the model (-2 log likelihood = 272.12, χ^2 = 6.56, *df* = 3, *p* = .087). The Wald statistics in the MLR model with 'no accent' as the reference category indicated that information packaging significantly predicted the choice between H* and 'no accent' (*b* = 1.1, Wald = 4.59, *df* = 1, *p* = .032).[3] The odds ratios (Exp(B) values in the SPSS output) showed that the odds of H* occurring in sentence-initial focus compared to sentence-initial topic was 3 times higher than that of 'no accent'.[4] Information packaging, however, did not significantly predict the choice between H*L and 'no accent' (*b* = 0.73, Wald = 2.4, *df* = 1, *p* = 1.21) as both patterns occurred more frequently in the topic condition than in the focus condition (Table 1). The Wald statistics in the MLR model with H*L as the reference category showed that there was no significant difference in the odds of H* and H*L being used in the focus condition compared to the topic condition. These results indicated that the use of 'no accent' differed in different conditions in sentence-initial position, in particular compared to the use of H*, but choice of accent type was similar across conditions.

Sentence-final nouns: The MLR modelling showed the model fitting was significantly improved after the variable information packaging was added to the model (- 2 log likelihood = 343.63, χ^2 = 32.97, *df* = 4, *p* < .001), indicating a significant overall relationship between information packaging and intonation in sentence-final nouns. The Wald statistics in the model with 'no accent' as the reference category showed that information packaging significantly predicted the choice between H*L and 'no accent' (*b* = 2.06, Wald = 17.82, *df* = 1, *p* < .001), the choice between !H*L and 'no accent' (*b* = 1.53, Wald = 15.44, *df* = 1, *p* < .001), and the choice between L*H and 'no accent' (*b* = 1.89, Wald = 23.1, *df* = 1, *p* < .001) in the focus condition compared to the topic condition. The odds ratios indicated that H*L, !H*L and L*H were 7.83 times, 4.6 times, and 6.61 times respectively more likely to occur than 'no accent' in the focus condition. The Wald statistics in the model with H*L as the reference category showed that there was no significant difference between H*L and !H*L (*b* = - 0.53, Wald = 1.65, *df* = 1, *p* = .199) and between H*L and L*H (*b* = - 0.17, Wald = 0.17, *df* = 1, *p* = .684). The odds of H*L, !H*L and L*H occurring in focus was thus similar to the odds of these accent

---

[3]The Wald statistic of a predictor or a predictor category is comparable to the t-statistic in a linear regression. It is the value of the regression coefficient (*b*) of the predictor (category) divided by its associated standard error.

[4]Odds are defined as the probability of an event occurring divided by the probability of an event not occurring. The odds ratio is the proportionate change in odds, calculated by dividing the odds after a unit change in the predictor by the odds before that change. It serves as an indicator of the change in odds resulting from a unit change in the predictor, similar to the *b* coefficient but is easier to interpret because it does not involve a logarithmic transformation. If the odds ratio is larger than 1, it indicates that as the predictor increases, the odds of the outcome occurring increases. If the odds ratio is smaller than 1, it indicates that as the predictor increases, the odds of the outcome occurring decreases (Field 2009).

types occurring in topic. These results confirmed that accent placement differed in different conditions in sentence-final position but not choice of accent type.

Taken together, the results show that the four- to five-year-olds realised the focused noun mostly with accentuation in both sentence positions, like adults. Further, they accented the sentence-initial focused noun mostly with H*L and H*, like adults. However, different from adults, they showed no preference for H*L over the other accent types (i.e. !H*L and L*H) in realising the sentence-final focused noun. As regards the realisation of the unfocused noun, the four- to five-year-olds realised the pre-focus noun (sentence-initial topic) mostly with H*L and H* and the post-focus noun (sentence-final topic) preferably with 'no accent', similar to adults. Further, when 'no accent' was used in sentence-initial position, this occurred more frequently in the unfocused nouns than in the focused nouns, as found in the production of adults.

Chen (in press) argued that the use of !H*L and L*H in addition to H*L in sentence-final focus was triggered by different kinds of motivations. In respect of the use of !H*L, adults also use !H*L in sentence-final focus but to a lesser extent than H*L. !H*L is not associated with the expression of newness vs. givenness in Dutch. Instead, it has the connotation that the speaker is no longer interested in further discussion on the current topic (Gussenhoven et al. 2003). Possibly, children have not acquired the meaning of !H*L at the age of four and five and consequently interpreted the instances of !H*L in sentence-final focus in adult speech as equally acceptable as instances of H*L. In contrast, the use of L*H might either come from children's need to seek confirmation from the experimenter on their response, in spite that they were put into charge in the game, or reflect how some young children habitually speak (i.e. ending sentences with a final rise).

5 Study 3: Seven- to Eight-Year-Olds

Study 3 is concerned with seven- to eight-year-olds' phonological focus-marking in full sentences (Chen in press).

5.1 Method

Twenty-three seven- and eight-year-olds were tested using the same method as in Study 2. Following the same data selection procedure as described in Section 4.1, full-sentence responses from 11 seven- and eight-year-olds (age range: 7;5 – 8;10, mean age 7;11) were intonationally transcribed.

5.1.1 Results and Discussion

The intonation patterns in 277 sentence-initial nouns and 279 sentence-final nouns were included for further analysis. As can be seen in Table 2, in sentence-

Table 2 Mean percentage distributions of the intonation patterns in the nouns (seven- to eight-year-olds) (adapted from Table 4 in Chen in press)

		H*	H*L	!H*L	L*H	OTHER [a]	no accent
initial	focused	22.7%	64.8%	n.a[b]	7.5%	3.5%	1.5%
	unfocused	27%	53.3%	n.a	6.6%	3%	10%
final	focused	4.2%	59.8%	15.1%	14%	0	6.9%
	unfocused	3.6%	29.9%	22.5 %	4.2%	0	39.8%

[a] The category 'OTHER' refers to infrequently used accent types.
[b] The accent type did not occur in the condition to which the cell corresponds.

initial position, the focused noun was nearly always accented, most frequently with H*L, followed by H*. The unfocused noun was mostly accented, most frequently with H* and H*L. When a sentence-initial noun was unaccented, this occurred almost only in the unfocused condition. In sentence-final position, the focused noun was mostly accented, most frequently with H*L, followed by !H*L, whereas the unfocused noun was frequently realised with 'no accent'. When the unfocused noun was realised with an accent, the most frequently used accent type was !H*L. These patterns suggested that information packaging mattered to accent placement and choice of intonation pattern in both sentence positions.

MLR modelling was performed at the significance level of 0.05 on the intonational patterns in the nouns in different sentence positions statistically. The independent variable was INFORMATION PACKAGING with two categories, topic (non-focus) and focus. The variable SPEAKER was used to define the subgroups of the data in the model. The dependent variable was the intonation in the nouns, consisting of four categories (H*, H*L, OTHER, and 'no accent') in sentence-initial position and four categories (H*L, !H*L, OTHER, and 'no accent') in sentence-final position. The reference category was either 'no accent' or H*L.

Sentence-initial nouns: The MRL modelling showed that the model fitting was significantly improved after the variable INFORMATION PACKAGING was added to the model (-2 log likelihood = 242.65, χ^2 = 14.03, *df* = 4, *p* = .007), indicating a significant overall relationship between information packaging and intonation in sentence-final nouns. The Wald statistics in the model with 'no accent' as the reference category indicated that INFORMATION PACKAGING significantly predicted the choice between H*L and 'no accent' (*b* = 2.36, Wald = 4.77, *df* = 1, *p* = .029). The odds of H*L occurring in sentence-initial focus compared to sentence-initial topic was 10.62 times higher than that of 'no accent'. The odds of H* occurring in sentence-initial focus compared to sentence-initial topic was similar to that of 'no accent' in that both patterns occurred more frequently in topic than in focus on average. The Wald statistics in the model with H*L as the reference category showed that INFORMATION PACKAGING significantly predicted the choice between H* and H*L (*b* = -.0.67, Wald = .37, *df* = 1, *p* = .012). The odds of H*L occurring in sentence-initial focus compared to sentence-initial topic was 1.95 times higher than that of H*, suggesting a preference for H*L in marking sentence-initial focus.

Sentence-final nouns: The MLR modelling showed that the model fitting was significantly improved after the variable INFORMATION PACKAGING was added to the model (-2 log likelihood = 286.28, χ^2 = 115.7 *df* = 3, $p < .001$), indicating a significant overall relationship between information packaging and intonation in sentence-final nouns. The Wald statistics in the model with 'no accent' as the reference category showed that INFORMATION PACKAGING significantly predicted the choice between !H*L and 'no accent' (b = 1.93, Wald = 31.68, *df* = 1, $p < .001$) and between H*L and 'no accent' (b = 3.32, Wald = 64.03, *df* = 1, $p < .001$). The odds of !H*L and H*L occurring in sentence-final focus relative to sentence-final topic was 6.89 times and 27,68 times respectively higher than that of 'no accent', suggesting a preference for accenting focus in sentence-final position. The Wald statistics in the model with H*L as the reference category showed that INFORMATION PACKAGING significantly predicted the choice between !H*L and H*L (b = -1.39, Wald = 10.74, *df* = 1, $p < .001$). The odds of H*L occurring in sentence-final focus relative to sentence-final topic was 4 times higher than that of !H*L, suggesting a preference for accenting sentence-final focus with H*L.

To sum up, like adults, the seven- to eight-year-olds realised the focused noun and the pre-focus noun predominantly with accentuation, and the post-focus noun mostly without accentuation. Further, the seven- to eight-year-olds showed an adult-like preference for H*L over !H*L in marking the focused noun in sentence-final position. This was a significant development compared to the four- to five-year-olds, who did not vary choice of accent type to mark sentence-final focus. Interestingly, the seven- to eight-year-olds also showed a preference for H*L in realising sentence-initial focused nouns, but not in realising sentence-initial unfocused nouns, unlike adults, who favoured H*L in both cases. Considering that adults varied the duration and pitch span of the H*L accent to distinguish focused sentence-initial nouns from unfocused sentence-initial nouns but the four- to five-year-olds did not do so, Chen (in press) interpreted the use of accent type as a useful strategy in the developmental stage when the seven- to eight-year-olds could only vary the pitch span of H*L for this purpose.

6 General Discussion

The results from the three studies show clearly that children acquire the use of accent placement and accent type to mark focus in a gradual fashion. The developmental path to adult-like phonological focus-marking is as follows. In the late two-word stage (at about the age of two), children appear to use accent type to mark focus, due to difficulty with unaccenting. More specifically, they associate non-downstepped accent types (e.g. H*L, H*) with the focused word and the downstepped H*L (and devoicing) with the unfocused word. In contrast, adults realise the unfocused word in two-word utterances without accentuation and they do not devoice. There seems to be no clear development between two and three. In the grammatical multiword stage (at the age of four or five), children are more skilled in unaccenting, and become adult-like in using accent placement to distinguish a focused word from an unfocused word. However, their choice of accent type in sentence-final focus is not yet adult-like. They show no preference for H*L over the other accent types (!H*L and L*H) in realising sentence-final focused nouns. As argued in Chen (in press), this lack of preference for H*L can be attributed to several factors, including the need of seeking confirmation and checking, the tendency to end a sentence with a final rise - both leading to the frequent use of L*H, and the influence of the relatively frequent use of !H*L in adults. Between four and eight, the use of accent type is further developed. More specifically, the seven- to eight-year-olds show the adult-like preference for H*L over !H*L and L*H in marking sentence-final focus. Further, they also use accent type to distinguish a focused word from an unfocused word in sentence-initial position. As adults do not use accent type for this purpose, the next possible development regarding the use of accent type is that children will stop using accent type to distinguish a focused noun from an unfocused noun in sentence-initial position. This development may take place when children can fully rely on the phonetic parameters for this purpose, like adults. Future work on older children's phonetic focus-marking is needed to establish the development after the age of eight.

The proposed developmental path to adult-like phonological focus-marking in Dutch may be applicable to children acquiring a language like Dutch (e.g. English and German), in which accent placement and accent type both play a significant role in marking focus. However, many languages are not like Dutch in this respect. For example, in Parisian French the shape of accent patterns plays no role in marking focus but phrasing does (Jun and Fougeron 2000). The focused constituent tends to form an independent accentual phrase and the post-focused sequence is merged into the same phrase. In tone languages like Mandarin Chinese, the shape of the pitch contour in a word is lexically determined, and focus is mainly marked by variations in the pitch range of the focused constituent and the post-focus sequence and the duration of the focused constituent (e.g. Xu 1999; Y Chen 2006). The question is then what kind of developmental path children acquiring languages like French and Mandarin Chinese will go through to become adult-like in intonational focus-marking. Future work can be directed to such cross-linguistic comparisons to shed light on language-specific acquisition challenges that children face in the process of acquiring intonational focus-marking.

References

Balog, Heather, and David Snow. 2007. The adaptation and application of relational and independent analyses for intonation production in young children. *Journal of Phonetics* 35: 118–133.

Baumann, Stefan, Johannes Becker, Martine Grice, Martine, and Doris Mücke. 2007. Tonal and articulatory marking of focus in German. In Jürgen Trouvain, and William J. Barry (eds.) *Proceedings of the 16th International Congress of Phonetic Sciences ICPhS*, 1029–1032. Saarbrücken, Germany.

Behrens, Heike, and Ulrike Gut. 2005. The relationship between prosodic and syntactic organization in early multiword speech. *Journal of Child Language* 32: 1–34.

Boersma, Paul. 2001. *Praat, a System for Doing Phonetics by Computer*. Glot International 5(9/10): 341–345.

Chafe, Wallace. 1974. Language and consciousness. *Language* 50(1): 111–133.

Chen, Aoju. in press. Tuning information packaging: intonational realisation of topic and focus in child Dutch. *Journal of Child Language*.

Chen, Aoju. 2007. Intonational realisation of topic and focus by Dutch-acquiring 4- to 5-year-olds. In Jürgen Trouvain, and William J. Barry, (eds.) *Proceedings of the 16th International Congress of Phonetic Sciences*, 1553–1556. Dudweiler: Pirrot GmbH.

Chen, Aoju, and Paula Fikkert. 2007a. Intonation of early two-word utterances in Dutch. In Jürgen Trouvain, and William J. Barry, (eds.) *Proceedings of the 16th International Congress of Phonetic Sciences*, 315–320. Dudweiler: Pirrot GmbH.

Chen, Aoju, and Paula Fikkert. 2007b. Dutch 3-year-olds' use of intonation in marking topic and focus. Poster presented at *Generative Approaches to Language Acquisition*. Barcelona, Spain. September 2007.

Chen, Yiya. 2006. Durational adjustment under corrective focus in Standard Chinese. *Journal of Phonetics* 34: 176–201.

Field, Andy. 2009. *Discovering Statistics Using SPSS* (3rd edition). London: Sage Publications.

Féry, Caroline. 2007. The fallacy of invariant phonological correlates of information structural options. In Caroline Féry, Gisbert Fanselow, and Manfred Krifka, (eds.) *Interdisciplinary Studies on Information Structure* (vol. 6), 161–184. Potsdam: Audiovisuelles Zentrum der Universität Postdam and GS Druck und Medien GmbH.

Fikkert, Paula. 1994. *On the Acquisition of Prosodic Structure*. The Hague: Holland Academic Graphics.

Frota, Sónia, and Marina Vigário. 2008. Early intonation in European Portuguese. Talk given at the *Third Conference on Tone and Intonation*, Lisbon, September 2008.

Gundel, Jeanette K. 1999. On different kinds of focus. In Petra Bosch, and Rob van de Sandt, (eds.) *Focus: Linguistic, Cognitive, and Computational Perspective*, 293–305. Cambridge: Cambridge University Press.

Gussenhoven, Carlos. 2005. Transcription of Dutch Intonation. In Sun-Ah Jun, (ed.) *Prosodic Typology and Transcription: A Unified Approach*, 118–145. Oxford: Oxford University Press.

Gussenhoven, Carlos, Toni Rietveld, Joop Kerkhoff, and Jaques Terken. 2003. *ToDI* (2nd edition). http://todi.let.kun.nl/ToDI/home.htm

Hanssen, Judith, Jörg Peters, and Carlos Gussenhoven. 2008. Prosodic Effect of Focus in Dutch Declaratives. In Plínio A. Barbosa, Sandra Madureira, and César Reis (eds.) *Proceedings of Speech Prosody 2008*, 609–612. Campinas, Brazil: Editora RG/CNPq.

Hornby, A. Peter, and Wilbur A. Hass. 1970. Use of contrastive stress by preschool children. *Journal of Speech and Hearing Research* 13: 359–399.

Horne, Merle. 1990. Accentual patterning in 'new' vs 'given' subjects in English. *Working Papers of Department of Linguistics at Lund University*, 81-97.

Jun, Sun-Ah, and Cécile Fougeron. 2000. A phonological model of French intonation. In Antonis Botinis, (ed.) *Intonation: Analysis, Modeling and Technology*, 209–242. Dordrecht: Kluwer Academic Publishers.

Ladd, D. Robert. 1980. *The Structure of Intonational Meaning: Evidence from English*. Bloomington: Indiana University Press.

Lambrecht, Knud. 1994. *Information Structure and Sentence Form: Topics, Focus, and the Representations of Discourse Referents*. Cambridge: Cambridge University Press.

Levelt, Clara. 1994. *On the Acquisition of a Place*. PhD dissertation. University of Leiden.

MacWhinney, Brian, and Elizabeth Bates. 1978. Sentential devices for conveying givenness and newness: A cross-cultural developmental study. *Journal of Verbal Learning and Verbal Behavior* 17: 539–558.

Müller, Anja, Barbara Höhle, Michaela Schmitz, and Jürgen Weissenborn. 2006. Focus-to-stress alignment in 4 to 5-year-old German-learning children. In Adriana Belletti, Elisa Bennati, Cristiano Chesi, Elisa DiDomenico and Ida Ferrari, (eds.) *Proceedings of GALA* 2005, 393–407. Cambridge UK: Cambridge Scholars Publishing.

Prieto, Pilar, and Maria del Mar Vanrell. 2007. Early intonational development in Catalan. In Jürgen Trouvain, and William J. Barry, (eds.) *Proceedings of the 16th International Congress of Phonetic Sciences*, 309–314. Dudweiler: Pirrot GmbH.

Xu, Yi. 1999. Effects of tone and focus on the formation and alignment of F0 contours. *Journal of Phonetics* 27: 55–105.

Wells, Bill, and John Local. 1993. The sense of an ending: a case of prosodic delay. *Clinical Linguistics and Phonetics* 7(1): 59–73.

Wells, Bill, Sue Peppé, and Nata Goulandris. 2004. Intonation development from five to thirteen. *Journal of Child Language* 31: 749–778.

Wieman, Leslie. 1976. Stress patterns in early child language. *Journal of Child Language* 3: 283–286.

A Phonetic Study of Intonation and Focus in Nłeʔkepmxcin (Thompson River Salish)

Karsten A. Koch

1 Introduction

This paper presents the results of a phonetic analysis of intonational properties in Nłeʔkepmxcin (Thompson River Salish), and their interaction with the discourse categories of focus and givenness. The present research is the first such study to be undertaken in any Salish language, and is based on all new data in the form of conversational recordings. Although Thompson Salish is a stress language (Egesdal 1984; Thompson and Thompson 1992), the Stress-Focus Correspondence (e.g. Reinhart 1995) manifested in stress languages like English (Selkirk 1995; Féry and Samek-Lodovici 2006, etc.) is not relevant for Nłeʔkepmxcin. An acoustic phonetic analysis indicates that:

1. focused information is not marked with additional prosodic prominence, and
2. given information is not marked with reduced prosodic prominence.

In Section 2, I give a brief overview of Nłeʔkepmxcin, including how narrowly focused and given material is marked morphosyntactically. Section 3 reviews the Stress-Focus Correspondence, as well as Destress-Given (Féry and Samek-Lodovici 2006, and many others), a constraint governing the deaccenting of old information. I also review the acoustic phonetic correlates of focal accent and given deaccentuation. In Section 4, I present the results of an experiment comparing the acoustic phonetic attributes of neutral wide focus utterances, on the one hand, with narrow focus utterances on the other. The Stress-Focus Correspondence Principle predicts that narrow foci are marked with additional prosodic prominence, while Destress-Given dictates that given material is reduced in prosodic prominence. However, we shall see that these hypotheses are not supported in Nłeʔkepmxcin. Rather, the experiment confirms the (previously impressionistic) hypothesis that Salish languages do not

K.A. Koch (✉)
Research Fellow, Zentrum für Allgemeine Sprachwissenschaft, Berlin, Germany
e-mail: karstenak@hotmail.com

S. Frota et al. (eds.), *Prosodic Categories: Production, Perception and Comprehension*, Studies in Natural Language and Linguistic Theory 82,
DOI 10.1007/978-94-007-0137-3_6,

mark discourse prominence with sentential accent (e.g. Benner 2006 on Sencothen Salish). Further implications are discussed in Section 5, and Section 6 concludes.

2 Background: Nłeʔkepmxcin

Nłeʔkepmxcin is a member of the Northern Interior branch of Salishan, a family of 23 languages (for overviews, see Thompson and Thompson 1992; Kinkade 1992; Kroeber 1999; Davis and Matthewson 2009). It is spoken in southwestern British Columbia, Canada. Like all remaining Salish languages, Nłeʔkepmxcin is critically endangered, with most fluent speakers in their 60's or older. The present study is based on a corpus of conversational recordings collected during fieldwork with two female speakers of the *ƛ̓q̓əmcín*, or Lytton, dialect. Both are bilingual, also being fluent speakers of English.

The phonemic inventory is presented in Table 1. I will use the orthography developed by Thompson and Thompson (1992, 1996) for examples throughout this paper.[1]

Table 1 Phonemic inventory (adapted from Thompson and Thompson 1992)

CONSONANTS	labial	alveo-lar	alveo-palatal	velar	uvular	pharyn-geal	glot-tal
Stops	p	t	-	k kʷ	q qʷ	-	ʔ
Ejectives	p̓	t’	-	k̓ k̓ʷ	q̓ q̓ʷ	-	-
Lateral Eject.	-	ƛ̓	-	-	-	-	-
Nasal	m	n	-	-	-	-	-
Glottalized	m̓	n̓	-	-	-	-	-
Affricates	-	c̣ [ts]	c [tʃ]	-	-	-	-
Ejective	-	c̓ [ts’]	-	-	-	-	-
Fricatives	-	ṣ [s]	s [ʃ]	x xʷ	x̣ x̣ʷ	-	h
Lateral	-	ł	-	-	-	-	-
Approximant	(w)	z	y [j]	w	-	ʕ ʕʷ	-
Lateral	-	l	-	-	-	-	-
Glottalized	(w̓)	z̓	y̓	w̓	-	ʕ’ ʕ’ʷ	-
Glott. Lateral	-	l’	-	-	-	-	-

VOWELS		front		central		back
high		i	ị			u
mid		e		ə	ə̣	o
low		-		a		-

[1]See Thompson and Thompson (1992) also for surface variation of vowels across contexts.

Nłeʔkepmxcin is a predicate-initial language. In typical wide focus contexts, this initial predicate is a verb like *kəntes* 'help' in (1), or a light verb (auxiliary) like the imperfective *ʔex* in (2). The basic word order, at least in the Lytton dialect that is the subject of the present paper, is verb-subject-object (VSO). Subjects are topical and often null (Gerdts 1988; Kinkade 1990), though in wide focus discourse contexts like (1) and (2) it is not uncommon for transitive subjects to be overtly expressed. Second position clitics (2CL), including evidentials, clause-typing morphology, and the ubiquitous situational deictic *xeʔ* in (1) and (2), follow the first prosodic word. Acute accents mark word-level stress, while '=' and '-' mark syntactic cliticization and affixation, respectively.[2]

(1)	V	[2CL]	S	O
	[kən-t-Ø-és=xeʔ		e=skíxzeʔ-kt	e=sínciʔ-kt]$_{FOC}$.
	help-TRANS-3O-3S=DEM		DET=mother-1PL.PS	D=younger.brother-1PL.PS

'[Our mother helped our brother]$_{FOC}$.'
(judgement in out-of-the-blue context: *'Our brother helped our mother.')

(2)	Aux [2CL]	V	S	O
	[ʔéx=xeʔ	c̓ax̣-t-Ø-és	ɬ=n-sxáy̓wi	e=swúx^{w}t]$_{FOC}$.
	IMPF=DEM	clean-TR-3O-3S	DET=1SG.POSS-husband	DET=snow

'[My husband was cleaning up the snow]$_{FOC}$.'

The sentences in (1) and (2) are from wide focus contexts, answering a question like 'What was going on?' Thus, the entire clause (CP, or Complementizer Phrase) is in focus. Because these utterances lack narrow foci or given information, they constitute the neutral focus case and are expected to carry the default stress pattern (e.g. Hayes and Lahiri 1991: 56; Selkirk 1995). I mark the focus with square brackets and the subscript 'FOC'. The pitch tracings and waveforms of focus neutral sentences like this show a gradual declination of F0 and intensity; example (2) is shown in Fig.1. Pitch accent peak F0 tends to occur early in the stressed vowel.

We have seen that, in instances of wide focus (e.g. 1), Nłeʔkepmxcin speakers employ a verb-initial structure. In fact, even narrow focus structures are predicate-initial in Thompson Salish.[3] There are two possible ways to

---

[2]The key to the abbreviations in the glosses is as follows: - = affix, = = clitic, Ø = phonologically null, 1 = 1st person, 2 = 2nd person, 3 = 3rd person, CLEFT = cleft predicate, COMP, C = complementizer, DEM = demonstrative, DET, D = determiner, EMPH = emphatic marker, EVID = evidential (*ek^{w}u* 'hearsay'), FOC = focus, IMPF = imperfective, NEG = negation predicate, NOM = nominalizer, OBJ, O = object, PL = plural, POSS, PS = possessive, Q = yes/no question marker, REFL = reflexive, SG = singular, SUBJ, SBJ, S = subject, SUBJ.GAP = transitive subject gap marker, TRANS, TR = transitive.

[3]Contrastive topics precede the matrix predicate. In these cases, however, the contrastive topic is set off in its own intonational phrase. I will not discuss contrastive topics further in this paper.

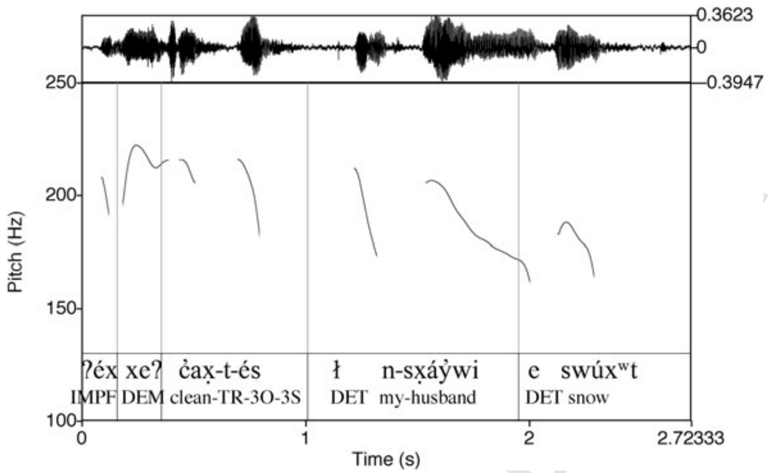


Fig. 1 Pitch tracing and waveform for (2): '[My husband was cleaning up the snow]$_{FOC}$'

mark narrow focus on a subject or object so that it is in the initial predicate (Kroeber 1997, 1999; Koch 2008a). The first is to use the focused bare noun as the initial matrix predicate, in what has been termed a 'nominal predicate construction' or NPC (Davis et al. 2004). Unfocused, or given, information follows in a residue clause, which is introduced by a complementizer and marked with subordinating morphology. In (3B), the focus is 'tea,' since it answers the object focus wh-question in (3A). The bare noun *tiy* in (3A) serves as base-generated, matrix predicate, and takes the residue clause *e nsx̣ʷox̣ʷst* 'that my wanting' as its subject. Subordination of the verb 'want' is marked by nominalization morphology in this case (see Kroeber 1997 for a thorough description of the paradigm of subordination morphology). It should be noted that nominal predicates are possible independent of focus context.

(3) A: *What do you want to drink?*
B: [tíy]$_{FOC}$ e=n=s=x̣ʷóx̣ʷ-st.
tea COMP=1SG.POSS=NOM=want-REFL
'I want [tea]$_{FOC}$.' (more literally: 'That my wanting is [tea]$_{FOC}$.')

The second way to mark narrow focus is through a cleft. The cleft predicate *ċe* or *ʔe* takes a focused Determiner Phrase (DP) as its first argument (thus maintaining the predicate-initial generalization). The focused DP is base-generated here. This is again followed by a residue clause introduced by a complementizer and containing the subordinated verb, which is given, or old, information. In (4), the focus is the DP *e Monik*, and the residue clause is *e wiktne* 'that I saw.'

(4) A: *Who did you see?*
B: c̓é=xeʔ [e=Moník]$_{FOC}$ e=wík-t-Ø-ne.
CLEFT=DEM DET=Monique COMP=see-TRANS-3OBJ-1SG.SUBJ
'I saw [Monique]$_{FOC}$.' (more literally: 'It was [Monique]$_{FOC}$ that I saw.')

Because subjects are usually definite DPs, subject focus is expressed by using clefts rather than nominal predicate constructions. Given, transitive verbs are marked with a special -*(e)mus* suffix indicating a subject gap in the residue clause (Kroeber 1997).

(5) A: *I heard that it was Fred who painted it.*
B: téʔe. c̓é [ł=Ross]$_{FOC}$ e=pínt-et-Ø-mus.
NEG. CLEFT DET= Ross COMP=paint-TRANS-3O-SUBJ.GAP
'No. It was [Ross]$_{FOC}$ that painted it.'

It is clear that focus is marked syntactically as well as morphologically in Nłeʔkepmxcin. I will not discuss the syntax of these constructions in greater detail, as my primary aim is to examine their intonation. For further discussion, see Kroeber (1997, 1999), Davis et al. (2004), and Koch (2007, 2008a, 2008b).

In the narrow focus structures in (3), (4) and (5), we see that the focus is the leftmost lexical item in the clause – that is, the focus bears the leftmost phrasal accent (the cleft predicate is a functional element and does not bear phrasal accent). Given material (typically the verb), on the other hand, is removed from this leftmost position and generated in a residue clause at the right edge of the utterance.[4] This contrasts with the wide focus structures in (1) and (2), where the verb is the leftmost lexical element and hence bears the initial phrasal accent. Arguments, on the other hand, follow the verb in these focus-neutral cases.

Strikingly, previous impressionistic observations have noted the lack of stress-focus effects in Salish:

> ...a common strategy when one wants to emphasize something in Sencothen is simply to make it the predicate. Based on the available data, it would seem that this syntactic strategy often reduces the need for prosodic strategies such as contrastive stress.
> (Benner 2006: 14, on Sencothen Salish)

This hypothesis will be tested experimentally in Section 4. The pitch tracings and waveforms of (3) to (5) are shown in the Fig. 2, to illustrate typical utterances.

The leftmost focus 'tea' is marked with a pitch accent, but not more prominently than left-edge verbs in neutral focus utterances; and the given verb *x̣ʷox̣ʷst* 'want' is also marked with a pitch accent. There is a gradual declination from left to right, similar to that in neutral focus utterances. Similarly, in Fig. 3, both the focus 'Monique' and the given verb *wiktne* are marked with pitch accents.

[4]Like in English clefts, given material (the residue clause) need not be overtly expressed.

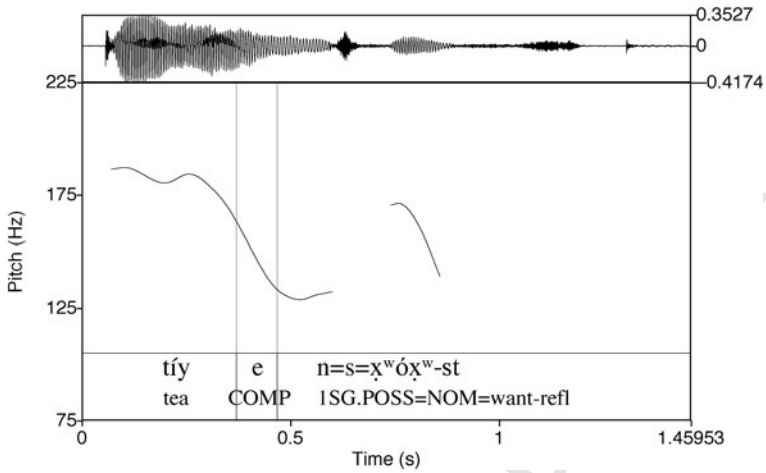


Fig. 2 Pitch tracing and waveform for (3): 'I want [tea]$_{FOC}$'

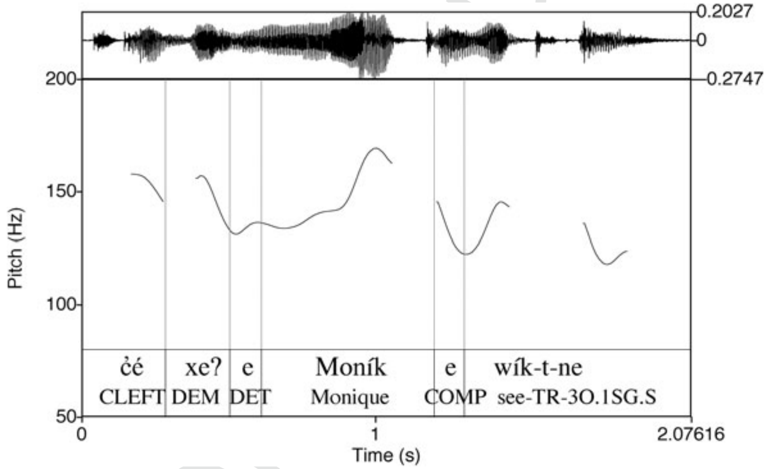


Fig. 3 Pitch tracing and waveform for (4): 'I see [Monique]$_{FOC}$'

Finally, the subject focus example in Fig. 4 is also characterized by gradual declination from left to right, and pitch accents on both the narrow focus 'Ross' and the given verb *pintetmus*.

In Section 4, I will present the results of an experiment which compares the acoustic phonetic correlates of these narrow focus utterances with the neutral focus cases like (1) and (2), and tests the hypothesis that the Stress-Focus Correspondence is not relevant for marking focus in Thompson Salish. First, though, I give some background on focus, givenness, and their acoustic phonetic correlates.

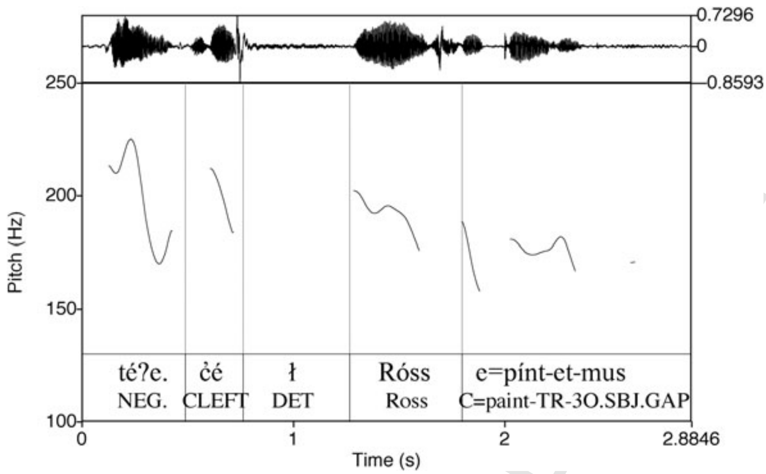


Fig. 4 Pitch tracing and waveform for (5): 'No. It was [Ross]$_{FOC}$ that painted it'

3 Background: Focus and Givenness

In this section, I will make clear how I define focus and givenness in this paper. After presenting the Stress-Focus Correspondence Principle and the DESTRESS-GIVEN constraint, I will then review previous research on the acoustic correlates of focal accent and given deaccenting.

3.1 Focus and the Stress-Focus Correspondence

In this study, I make the standard assumption that focus is a syntactic category, identified by f(ocus)-marks, or a [FOCUS] feature (Jackendoff 1972; Selkirk 1995; etc.; Rooth 1992 introduces a syntactic focus operator). This syntactic focus feature mediates between the phonological expression of focus and its semantic interpretation. Because focus is an interface phenomena, it is widely assumed to be visible to phonological constraints like STRESS-FOCUS (Féry and Samek-Lodovici 2006), thus crossing from being a purely syntactic category to a prosodic category as well (Reinhart 2006).

Focus can be identified in different ways. It is classically diagnosed as the answer to a wh-question (Halliday 1967; Jackendoff 1972; Selkirk 1995, and many others). Wh-questions do not need to be overt in the discourse; information that answers an implicit wh-question is also focused (van Kuppevelt 1994). However, since implicit questions could not be determined *a priori*, they were not employed as a method for identifying focus in this study.

Bart in (6) is a focus, since this DP answers the overt *who* question.

(6) A: Who cooked dinner?
B: [Bart]$_{FOC}$ cooked dinner.

Focus can also stand in contrast to a previous (part of an) utterance, rather than wh-question:

(7) A: I heard Janice found some mushrooms.
B: No, [Kelly]$_{FOC}$ found some mushrooms.

Focus sequences thus involve dual and symmetric frames in which only one element differs. The background is shared between both configurations. In (7), this frame is *x found some mushrooms*, and *Janice* and *Kelly* are the contrastive foci in a type of anaphoric relationship (Rochemont 1986; Rooth 1992: 80; Féry and Samek-Lodovici 2006: 135).

In the present study, focus was identified in discourse exchanges of the type in (6) and (7).

Prosodically, it has been widely observed that focused constituents in stress languages bear the nuclear pitch accent. Previous proposals refer to both the marking of focus (8a-c), and to the interpretation of focus (8d-e).

(8) Proposals on the marking of focus
a. Focus: A Focus-marked phrase contains an accent. (Schwarzschild 1999: 173)
b. Focus-Prominence: Focus needs to be maximally prominent. A prosodic category C that contains a focused constituent is the head of the smallest prosodic unit containing C. (Truckenbrodt 1995)
c. Stress-Focus: a focused phrase has the highest prosodic prominence in its focus domain. (Féry and Samek-Lodovici 2006: 135-136)

Proposals on the interpretation of focus
d. Basic Focus Rule: An accented word is F(ocus)-marked. (Selkirk 1995: 555)
e. Stress-Focus Correspondence Principle: The focus of a clause is a(ny) constituent containing the main stress of the intonational phrase, as determined by the stress rule. (Reinhart 1995: 62)

What all of the proposals above have in common is the alignment of focus with the nuclear pitch accent, the prosodic head of the intonational phrase. While the proposals cited above were formulated largely based on English, the correspondence between focus and sentential stress is claimed to be a universal feature of stress languages (and shown to be so in many studies on other Indo-European languages):

> Intonation languages use pitch accents as the principal means of focusing. Most intonation languages use the H*L falling tone as a pitch accent to mark focus, where

> the * following the H tone tone signals that the tone on the accented syllable is high. (Hartmann 2007: 225)

3.2 Givenness and DESTRESS-GIVEN

Roughly speaking, given material is old information that is already in the Common Ground of a conversation. In identifying given material, I follow Schwarzschild (1999), who proposes that given constituents are entailed by prior discourse via coreference (for entities) or existential F-closure.

A common property of stress languages is the deaccenting of given information, particularly after the nuclear pitch accent.

> The postfocal contour is deaccented, due to the fact that there are no more accent targets following the focus. Thus, the pitch range, which is expanded on the focus constituent, is compressed post-focally. (Hartmann 2007: 225-226)

This generalization is described in the constraint DESTRESS-GIVEN:

(9) DESTRESS-GIVEN: A given phrase is prosodically non-prominent. (Féry and Samek-Lodovici 2006: 135-6)

Deaccenting of given information does not appear to be a universal property of stress languages (Gumperz 1982 on Indian and Caribbean English; Ladd 1996 on Italian, Romanian, and a general overview of languages that lack deaccenting).

3.3 Acoustic Correlates of Focal Accent and Givenness Deaccenting

In Section 2, we observed that the effect of narrow focus in Nłeʔkepmxcin is to reverse the linear position of verb and argument Table 2.

In the acoustic phonetic study in Section 4, I exploit this change in linear position to compare the intonation contours of neutral focus utterances on the one hand, with narrow focus utterances on the other. All else being equal, we expect the leftmost accent in narrow focus utterances (the focus) to be more prominent than the leftmost accent in the neutral focus cases (the verb). This follows from the Stress-Focus Correspondence Principle. On the right edge, we expect the given material in narrow focus cases (typically the verb) to bear lesser prosodic prominence than the rightmost accent in the focus-neutral cases (an argument of the verb). This follows from DETRESS-GIVEN.

Table 2 Linear order in wide focus (neutral focus) and narrow focus contexts

Wide focus (neutral focus):	VERB	…	ARGUMENTS
Narrow subject/object focus:	FOCUS	…	VERB

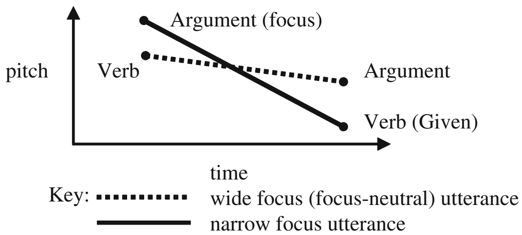


Fig. 5 Predicted intonation contours of wide and narrow focus utterances

In Fig. 5, we see three measurable ways that the intonation contours of the two utterance types are expected to differ. First, the focus at the left edge is realized with a higher degree of prosodic prominence. Secondly, given material at the right edge is realized with a lesser degree of prosodic prominence. Finally, the declination line between these two points is steeper in the narrow focus case. I will be measuring declination from peak to peak, or 'topline' declination (e.g. 't Hart et al. 1990; Strik and Boves 1995).

Now we are in a position to ask what the acoustic phonetic correlates of this type of focal accenting and givenness deaccenting are. Classic markers of prominence in stress languages are pitch (fundamental frequency, or F0), duration, loudness (intensity, often informally referred to as amplitude) and vowel quality (Fry 1958; Lieberman 1967, etc.). F0 has been the most consistently linked acoustic cue for pitch accents (see Gussenhoven 2004 for thorough discussion), particularly cues related to maximum F0. As an anonymous reviewer points out, low tone marking can also be relevant (e.g. Chen and Gussenhoven 2008 on tonal modification in Standard Chinese, Edwards 1954 and Muehlbauer 2005 on Nêhiyawêwin (Plains Cree)); however, since there was no prior evidence (either impressionistic or in the literature) for low tone marking of focus in Thompson Salish, nor any evidence in the inspected pitch tracings, I do not report on F0 minimum in the study in Section 4.[5]

Measurements will be made of the accented vowels corresponding to the elements of Fig. 5: the left edge, the right edge, as well as any intermediary accented vowels. I will briefly review Stress-Focus Correspondence studies of other languages, in which the researchers performed measurements as in Fig. 5 – that is, they compared the acoustic correlates of stressed vowels in neutral focus utterances, on the one hand, with their acoustic correlates when there was a narrow focus in initial position, on the other hand.

After reviewing these studies, I will take the most conservative results; assuming universal cognitive perceptual abilities, these conservative results thus represent a universal minimal difference required to make focus marking

[5]It may be added that the study in section four measures F0 range and timing of F0 peaks, which would be affected by any low tone marking of focus. As we shall see, however, the results provide no evidence for the role of low tones in pitch accent and focus marking.

acoustically perceptible to a hearer (see e.g. ’t Hart 1981 for evidence along these lines for Dutch), and we would expect to find similar acoustic marking in Thompson River Salish. If we do not find even the most minimal differences that are reported for other languages, then this will be an interesting result.

For accented vowels at the left edge, narrow focus has been found to induce +2 semitones greater F0 peak (Shue et al. 2007, on American English), +2 semitones greater F0 peak (Grønnum 1998: 142-3, Fig. 6, on Danish), +2.3 semitones greater F0 peak (Gårding 1998: 125, Fig. 4, on Swedish), and +2.0 semitones greater F0 peak (Suomi et al. 2003, on Finnish).

For F0 excursion at the left edge, previous studies have found increases in F0 range of +3.6 semitones (Eady and Cooper 1986, on American English), +3.4 semitones (Suomi et al 2003:122, on Finnish), and +2.3 semitones greater F0 range combined with a +1.5 semitone greater F0 peak (Eady et al. 1986, on American English).

Timing of the F0 peak was found to be earlier by 27% of the syllable duration by Eady and Cooper (1986) and Eady et al. (1986); but Shue et al. (2007), also studying American English, found that F0 peaks in the narrow focus condition occurred about 100 ms later.

Finally, in terms of duration, previous researchers have noted increases in duration of +31.2% (Eady and Cooper 1986: 407, Table II, on American English), +34.4% (Eady et al. 1986: 244, Table 3, on American English),

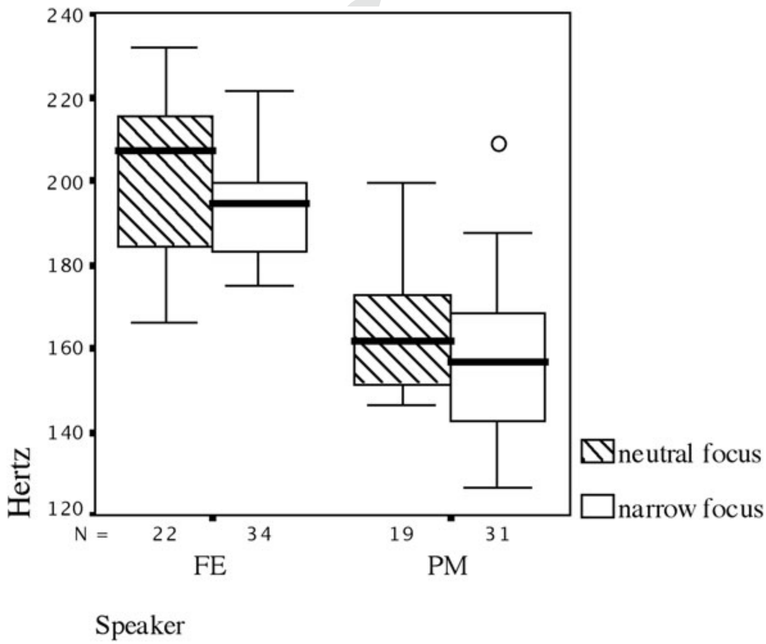


Fig. 6 Maximum left edge F0 by speaker and focus type

Table 3 Hypotheses for acoustic correlates of narrow focal accent at the left edge (μ_1 = narrow focus, μ_2 = neutral focus condition)

F0 peak	$\mu_1 - \mu_2 \geq 2$ semitones
F0 range	$\mu_1 - \mu_2 \geq 3.6$ semitones
F0 peak timing	$\mu_1 - \mu_2 \geq 27\%$ variation

+38%–41% (Cooper et al. 1985: 2146, Table III, on American English), +12.1% (Botinis 1998: 302, on Greek), and +11%–37% (Suomi et al. 2003:122, on Finnish).

Given these findings, we can take the most conservative figures, resulting in the following null hypotheses Table 3. F0 peaks are expected to increase by a minimum of +2 semitones (e.g. Shue et al. 2007). In the absence of greater F0 peaks, we may find greater F0 range, which is expected to increase +3.6 semitones (e.g. Eady and Cooper 1986). Timing of the pitch peak is expected to vary by at least 27% (e.g. Eady et al. 1986). Durational estimates vary considerably, so I will simply test the standard null hypothesis. None of these studies provided information on intensity differences induced by narrow focus at the left edge.

At the right edge, deaccentuation of given material has been found to result in lower F0 peaks of –3.5 semitones (Shue et al. 2007: 2627, Fig. 2, on American English), –3.5 semitones (Astruc and Prieto 2006: adapted from Fig. 1, on Catalan), –6.65 semitones (Okobi 2006, calculated from Appendix B, on American English), –1.4 to –2.7 semitones (Eady and Cooper 1986: 407-408, Table III, on American English), –6 semitones (Benkirane 1998: 351-352, 356, on Moroccan Western Arabic), –4.0 semitones (Suomi et al. 2003, on Finnish), and –3.5 to –5.5 semitones (Butcher and Harrington 2003, Fig. 1). F0 range has been found to decrease by –4.1 semitones (Suomi et al. 2003, on Finnish).

For peak intensity, previous research has noted decreases of –2.9 dB (Sluijter and van Heuven 1996b: 2475, Table II, on Dutch), –5 dB (Sluijter and van Heuven 1996a: 3, est. from Figure I, on American English), –4 dB (Astruc and Prieto 2006: adapted from Fig. 1, on Catalan), and –5.36 dB (Okobi 2006, calculated from Appendix B, on American English).

Finally, previous studies have uncovered durational decreases of –6.1% to –11.9% (Sluijter and van Heuven 1996b: 2475, Table II, on Dutch), –22.3% (Okobi 2006, calculated from Appendix B, on American English), –17.3% to –19.6% (Sluijter and van Heuven 1996a: Table I, on American English), –11% (Astruc and Prieto 2006: adapted from Fig. 1, on Catalan), –16% to –23% (Turk and White 1999, on Scottish English), 0% (Eady et al. 1986: experiment 1, on American English), –8.3% (Eady et al. 1986: 244, experiment 2, Table 3, on American English), –16.7% (Cooper et al. 1985: experiment 1, on American English), 0% (Cooper et al. 1985: 2145, experiment 2, Table IV, on American English), and 0% (Suomi et al. 2003: 120, 122, on Finnish).

These findings give us the following hypotheses regarding acoustic correlates of deaccenting Table 4. Given material is expected to have –3.5 semitones lesser F0 peak (e.g. Shue et al. 2007). In the absence of many figures for F0 range, we

Table 4 Hypotheses for acoustic correlates of deaccenting given material at the right edge (μ_1 = narrow focus, μ_2 = neutral focus condition)

F0 peak	$\mu_1 - \mu_2 \leq -3.5$ semitones
F0 range	$\mu_1 - \mu_2 \leq -3.6$ semitones
Peak intensity	$\mu_1 - \mu_2 \leq -3$ dB
Vowel duration	$\mu_1 - \mu_2 \leq -0.06*\mu_2$

can adopt the left edge value of Eady and Cooper (1986), whereby we'd expect at least –3.6 semitones lesser F0 range (this being a more conservative estimate than Suomi et al. 2003 found for the right edge). Peak intensity is expected to decrease by –3 dB or more (Sluijter and van Heuven 1996b). Estimates of vowel duration range considerably; we can adopt the –6% value from Sluijter and van Heuven (1996b) as a conservative estimate.

Finally, few studies have addressed the amount of declination to expect between left and right edge accents in neutral versus narrow focus cases. From Table III in Eady and Cooper (1986: 407), we can calculate this declination to be –4.1 semitones in neutral focus utterances, and –6.3 semitones when there is a narrow focus at the left, a difference of –2.2 semitones. In the absence of clear predictions for declination between left and right peaks, I will test the standard null hypothesis.

4 Experiment: Neutral Versus Narrow Focus

In this section, I report the results of a detailed acoustic phonetic study of different focus types in Thompson Salish; neutral, wide focus on the one hand, and narrow object or subject focus on the other.

4.1 Subjects

The language data was collected from two female speakers of Nłeʔkepmxcin in their late 60's (FE and PM). Both are speakers of the Lytton dialect, and fluently bilingual in English.

4.2 Method

Different instances of focus were identified from a corpus of conversational recordings made over a 20 month period of fieldwork. Field recordings were made using a Marantz PMD 670, 671 or 660 digital audio recorder. Each consultant was recorded on a separate channel using a Countrymax Isomax EMW Lavalier lapel microphone.

Narrow focus was identified according to the criteria noted in Section 3.1. Wide focus, or focus-neutral, utterances started a discourse or answered a wide-focus question like 'What happened?' To account for declination effects, only utterances which were completed in a single breath group were entered into the phonetic analysis. In addition, utterances in both conditions occurred at the start of the speaker's discourse turn, to control for effects of global utterance position within a larger discourse unit. These criteria resulted in 64 cases of neutral focus being identified in the corpus, to be compared with narrow subject (56 cases) and object (54 cases), to determine if and how the acoustic signal differed.

For each utterance, stressed vowels were identified in Praat (Boersma and Weenink 2007). Of primary interest were the vowels of the left and right edge. The first lexical stressed vowel at the left edge is the verb in the default case, and the focused NP in the narrow focus utterances. These were the vowels that were compared when testing the hypothesis that narrowly focused items carry greater acoustic prominence. At the right edge, the rightmost stressed vowel in the default case is typically an argument, while in the focus cases it is given information in the cleft clause. These were the vowels that were compared when testing the hypothesis that given information is reduced in acoustic prominence. In addition, other stressed vowels throughout the utterance were identified, in order to provide a better overall picture of the declination contour, as well as a better account of variability.

Utterance length was also identified. For both individual vowels and entire utterances, a variety of acoustic measurements were then made by using automated scripts in Praat. Pitch measurements of primary interest were the maximum and minimum F0, and the timing of the F0 peak (expressed as a percentage of the vowel duration). Where the Praat algorithm mismeasured F0, measurements were done by hand via visual inspection of the waveform, and automated measurements were disregarded. The average and maximum intensity (deciBels) was also recorded, as was target vowel and utterance duration (milliseconds).

To ensure that the utterances and the left and right stressed vowels of interest in the two conditions were comparable, with respect to their position in the utterance and breath group, several controls were implemented. Since, as noted by an anonymous reviewer, sentence length can have an effect on F0 (e.g. Swerts et al. 1996; Shih 1997, etc.), as can the position of the target syllable within the utterance, it was important to carry out these checks. I briefly review the results of these controls now.

The left edge: general comparisons. The utterances in the two groups were similar. Mean utterance duration was 2.33 seconds in the neutral focus cases ($n=41$, $SD=0.83$), and 2.48 seconds in the narrow focus sentences ($n=65$, $SD=0.65$), a non-significant difference ($t=0.898$, $df=104$, $p>0.3$). The stressed vowels to be compared were an average of 2.41 syllables from the left edge in the default case ($n=41$, $SD=2.76$), and 2.48 syllables from the left in the narrow focus case ($n=65$, $SD=1.92$). Again, this difference was non-significant

($t = 0.137$, $df = 104$, $p > 0.8$). These controls suggest that any intonational differences between the two utterance types is not likely to be due to declination effects, but rather will reflect other factors (such as information structure).

The right edge: general comparisons. For neutral focus cases, utterance length was an average 2.27 seconds ($n = 39$, $SD = 0.82$ sec), while for narrow focus utterances it was 2.46 seconds ($n = 64$, $SD = 0.76$ sec). These means were not significantly different ($t = 1.143$, $df = 102$, $p > 0.25$). The rightmost stressed vowels that were measured were an average of 0.54 syllables from the right in the neutral focus sentences ($n = 39$, $SD = 0.75$), and 1.08 syllables from the right in narrow focus utterances ($n = 64$, $SD = 2.13$), a non-significant difference ($t = 1.521$, $df = 102$, $p > 0.1$). Thus, any differences uncovered between the two focus conditions are unlikely to be due to effects of declination or utterance-final lengthening, but instead due to other factors like information structure.

4.3 Statistical Analysis

Results were analyzed for means (M), standard deviations (SD), and statistical significance. Numbers of observations (n) and degrees of freedom (df) are also reported where relevant.

The null hypotheses are as follows:

(10) Null hypotheses
a. Narrowly focused items attract additional prosodic prominence: $\mu_1 \geq \mu_2$
b. Given material is reduced in prosodic prominence: $\mu_1 \leq \mu_2$

The experimental hypotheses to be tested are as follows:

(11) Experimental hypotheses
a. Narrowly focused items in Salish do not attract additional prosodic prominence: $\mu_1 - \mu_2 = 0$
b. Given material in Salish is not reduced in prosodic prominence: $\mu_1 - \mu_2 = 0$

The use of a null hypothesis where two means are expected to be unequal, and an experimental hypothesis where the means are expected to be equal, is less common than the reverse situation. However, it should be made clear that the statistical model does not specify a null hypothesis of $\mu_1 - \mu_2 = 0$. Null hypotheses can perfectly well 'specify outcomes rather than absence of an effect' (Keppel and Wickens 2004: 72), and an honest examination of the focus marking literature presents just such a case. The overwhelming result of research into focus marking in stress languages is that focus is marked by additional prosodic prominence; this is the null hypothesis, and its rejection would be an interesting result. A significant result allows us to reject the null; when the null is set up like in (10), this allows us to reject the stress-focus and destress-given hypotheses.

Thus, where possible, I specify a null hypothesis based on stress-focus generalizations in the literature, as discussed in Section 3.3. As noted in that section, I adopt conservative values, since these will make a significant result less likely and therefore more robust. I employ the t-test since it easily allows the null hypothesis to be set to one which anticipates a difference between means.

This methodology risks missing the possibility that prominence can be identified only by looking at some combination of F0, amplitude and duration, rather than values of individual variables. A complex multivariate model could get at this possibility, but is well beyond the scope of the present study. All of the studies reviewed in Section 3.3 also examine acoustic variables individually (and often only one or two acoustic variables). Thus, the present statistical design is broadly comparable to other research in the field (though more comprehensive than most in the range of variables examined and reported on).

Planned comparisons of the means were carried out using independent sample t-tests for each variable (using pooled variances). Due to the number of comparisons performed (30 comparisons are reported on), a p-value of 0.001 was chosen for significance, to avoid an inflated experiment-wise error rate. With $p = 0.001$, the experiment-wise error rate is limited to 0.03, close to the standard value of 0.05. To indicate trends in the data, however, I mark results at three levels: $p < 0.05$ and $p < 0.01$ trend in the expected direction without achieving significance (indicated with * and ** respectively), while $p < 0.001$ is the actual significance level (indicated by ***). I present the results for each speaker separately, since they have differing F0 ranges.

4.4 Results: The Leftmost Lexical Stress

The left edge: pitch (F0). Complete results are reported in Tables 5 and 6. For both speakers, narrowly focused NPs actually had, on average, a lower F0 peak, and similar peak timings and F0 ranges as in the neutral focus condition.[6]

Figure 6 shows the results for F0 peaks. Under the null hypothesis, narrowly focused items were expected to carry an F0 peak that was at least 2 semitones greater than the left edge verbs in the default focus cases. This hypothesis was not supported. Independent sample t-tests for both FE ($t = -9.00$, $df = 54$, $p < 0.001$) and PM ($t = -5.27$, $df = 48$, $p < 0.001$) were significant, allowing us to reject the null hypothesis that Thompson Salish speakers mark narrowly focused items in left edge clefts with greater F0 peaks.

[6]The boxplots (or 'box and whisker' plots) throughout this section display the data separated by speaker and focus type. The dark line in the box represents the median value, and the box show the interquartile range (the middle 50% of the data points). The whiskers in either direction represent values 1.5 times the interquartile range. Outliers are indicated as open circles or stars beyond this range.

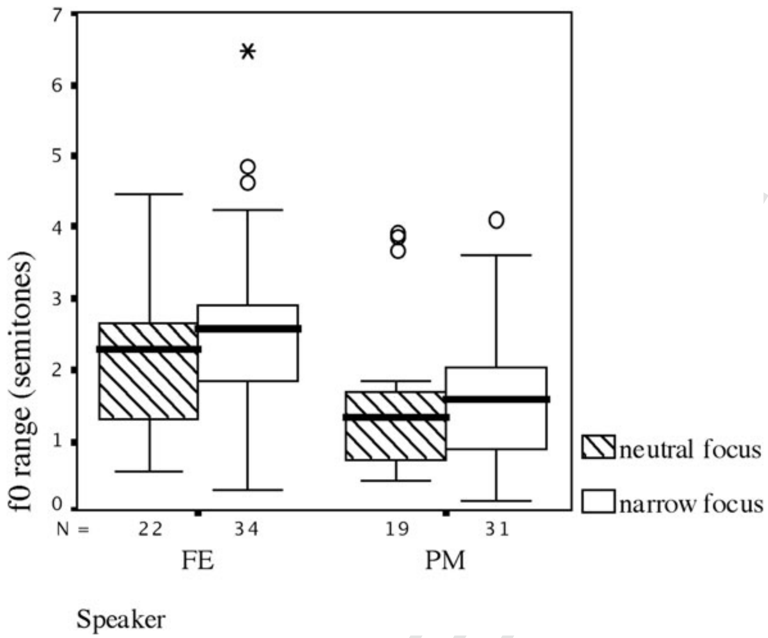


Fig. 7 Left edge F0 range (semitones) by speaker and focus type

The size of the F0 range, in semitones, is shown in Fig. 7. Based on previous research, the null hypothesis was that narrowly focused items at the left edge would show a greater pitch excursion of at least 3.6 semitones.

This null hypothesis was not supported. Neither speaker marked narrowly focused DPs with greater pitch excursions. Independent sample t-tests for both FE ($t = -9.89$, $df = 54$, $p < 0.001$) and PM ($t = -12.02$, $df = 48$, $p < 0.001$) were significant, allowing us to reject the null hypothesis that Thompson Salish speakers mark narrowly focused items with a 3.6 semitone greater F0 range.

The timing of F0 peaks is summarized in Fig. 8. Previous research suggests that, as a percentage of total vowel duration, narrow left edge focus may be marked by earlier pitch peaks of as little as 27% (Eady and Cooper 1986). In the present study, lexical items at the left edge were similarly marked in terms of timing of F0 peaks in both focus conditions, differing on average by only 4% for FE and 16% for PM. Because the data were not normally distributed (most F0 peaks occurred near the start of the vowel) and did not have equal numbers across conditions, I did not perform t-tests for this variable.

The left edge: intensity (dB). Review of previous studies, which tend to concentrate on F0, did not reveal any predicted differences for this variable. Thus, I tested the standard null hypothesis that the difference between means is 0. Independent sample t-tests failed to find a significant difference between maximum intensity (Fig. 9) in the neutral and narrow focus conditions, for both

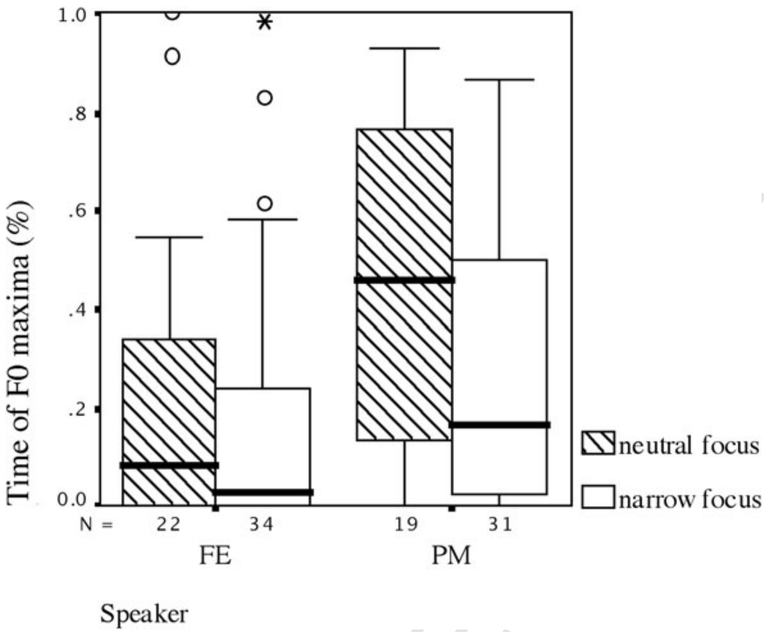


Fig. 8 Left edge time of F0 peak as a percentage of vowel duration

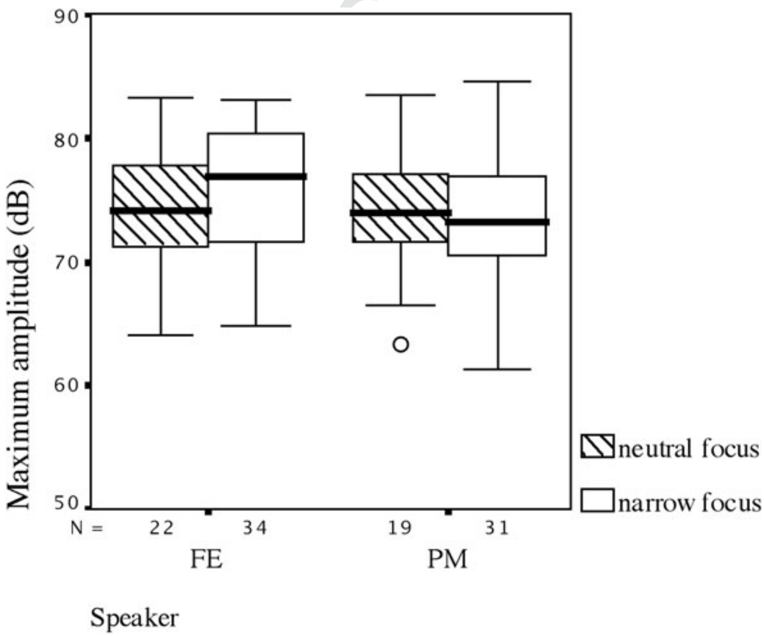


Fig. 9 Left edge maximum amplitude (dB)

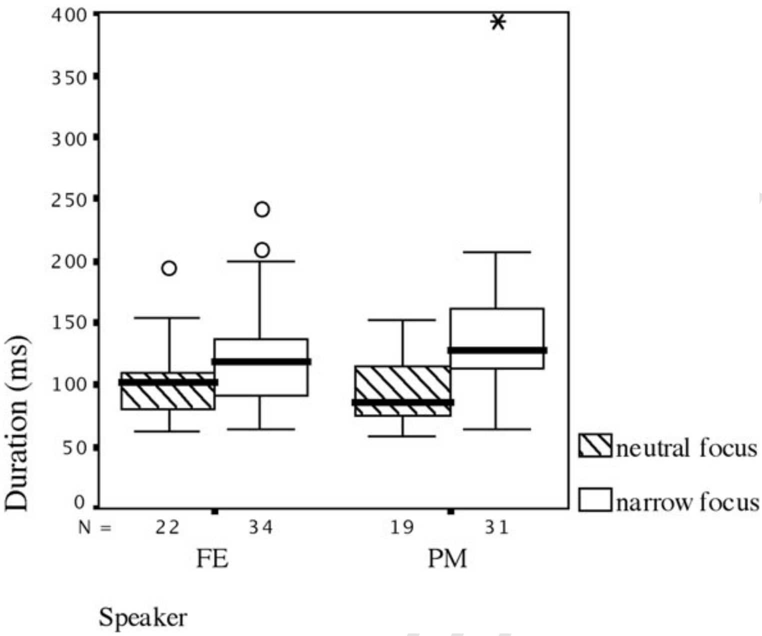


Fig. 10 Left edge vowel duration (ms) by speaker and focus type

FE ($t = -1.347$, $df = 54$, $p > 0.1$) and PM ($t = 0.586$, $df = 48$, $p > 0.5$). Results are similar for average vowel intensity: for neither FE ($t = -1.208$, $df = 54$, $p > 0.2$) nor PM ($t = 0.919$, $df = 48$, $p > 0.3$) were differences between the two conditions significant.

The left edge: duration (ms). Previous studies that have examined this correlate of focal accent varied considerably in their findings, though focal accent did usually increase duration (see Section 3.3). However, because of the lack of ample precedent, I tested the standard null hypothesis, that the two focus types did not differ in vowel duration of the leftmost stressed lexical vowel. Results (Fig. 10) were nor significant for either speaker ($t = -2.025$, $df = 54$, $p = 0.05$, for FE; $t = -3.365$, $df = 48$, $p = 0.002$, for PM).

The left edge: summary. Narrowly focused items in left edge clefts were not marked by increased pitch, intensity or duration. Results for each speaker are summarized in Tables 5 and 6.

4.5 Results: The Right Edge

The right edge: pitch (F0). Previous studies have found that deaccented material has F0 peaks that are 3.5 semitones or more lower than vowels carrying the nuclear pitch accent in the same position (e.g. Shue et al. 2007).

Table 5 FE leftmost lexical stress: summary of acoustic cues, and t-test results

Measure	Focus	Mean (SD)	*n*	Null Hypoth.	*t*	*p*	*df*
–	–	–	–	–	–	–	–
Max F0 (Hz)	Neutral	202.5 (16.9)	22	$\mu_1 - \mu_2 \geq 2$	−9.00	***<0.001	54
–	Narrow	193.8 (11.1)	34	semitones	–	–	–
–	–	–	–	–	–	–	–
F0 excursion (semitones)	Neutral	2.18 (1.03)	22	$\mu_1 - \mu_2 \geq 3.6$	−9.89	***<0.001	54
	Narrow	2.53 (1.30)	34	–	–	–	–
–	–	–	–	–	–	–	–
Time of F0 peak (%)	Neutral	21.0 (30.0)	22	$\mu_1 - \mu_2 \geq 0.27$	–	–	–
	Narrow	17.0 (27.0)	34	–	–	–	–
–	–	–	–	–	–	–	–
Maximum intens. (dB)	Neutral	74.2 (4.64)	22	$\mu_1 - \mu_2 = 0$ db	−1.347	>0.1	54
	Narrow	76.0 (4.97)	34	–	–	–	–
–	–	–	–	–	–	–	–
Average intens. (dB)	Neutral	72.2 (4.86)	22	$\mu_1 - \mu_2 = 0$ db	−1.208	>0.2	54
	Narrow	73.9 (5.29)	34	–	–	–	–
–	–	–	–	–	–	–	–
V duration (ms)	Neutral	103.4 (30.2)	22	$\mu_1 - \mu_2 = 0$ ms	−2.025	*0.05	54
	Narrow	123.7 (40.2)	34	–	–	–	–

Key: SD = standard deviation, n = number of observations, p = probability, df = degrees of freedom, μ = mean, *=trending significant at $p<0.05$, **= trending significant at $p<0.01$, ***=significant at $p<0.001$ [note: *** corresponds to a p-value of $p = 0.03$ after correcting for experiment-wise error using the Bonferroni procedure]

Table 6 PM leftmost lexical stress: summary of acoustic cues, and t-test results

Measure	Focus	Mean (SD)	*n*	Null Hypoth.	*t*	*p*	*df*
–	–	–	–	–	–	–	–
Max F0 (Hz)	Neutral	164.8 (15.0)	19	$\mu_1 - \mu_2 \geq 2$	−5.27	***<0.001	48
	Narrow	157.9 (19.1)	31	semitones	–	–	–
–	–	–	–	–	–	–	–
F0 excursion (semitones)	Neutral	1.55 (1.10)	19	$\mu_1 - \mu_2 \geq 3.6$	−12.02	***<0.001	48
	Narrow	1.64 (0.94)	31	–	–	–	–
–	–	–	–	–	–	–	–
Time of F0 peak (%)	Neutral	45 (33)	19	$\mu_1 - \mu_2 \geq 0.27$	–	–	–
	Narrow	29 (29)	31	–	–	–	–
–	–	–	–	–	–	–	–
Maximum intens. (dB)	Neutral	74.4 (4.90)	19	$\mu_1 - \mu_2 = 0$ db	0.586	>0.5	48
	Narrow	73.5 (5.58)	31	–	–	–	–
–	–	–	–	–	–	–	–
Average intens. (dB)	Neutral	72.3 (4.75)	19	$\mu_1 - \mu_2 = 0$ db	0.919	>0.3	48
	Narrow	70.8 (6.08)	31	–	–	–	–
–	–	–	–	–	–	–	–
V duration (ms)	Neutral	96.4 (27.5)	19	$\mu_1 - \mu_2 = 0$ ms	−3.365	**0.002	48
	Narrow	145.2 (59.3)	31	–	–	–	–

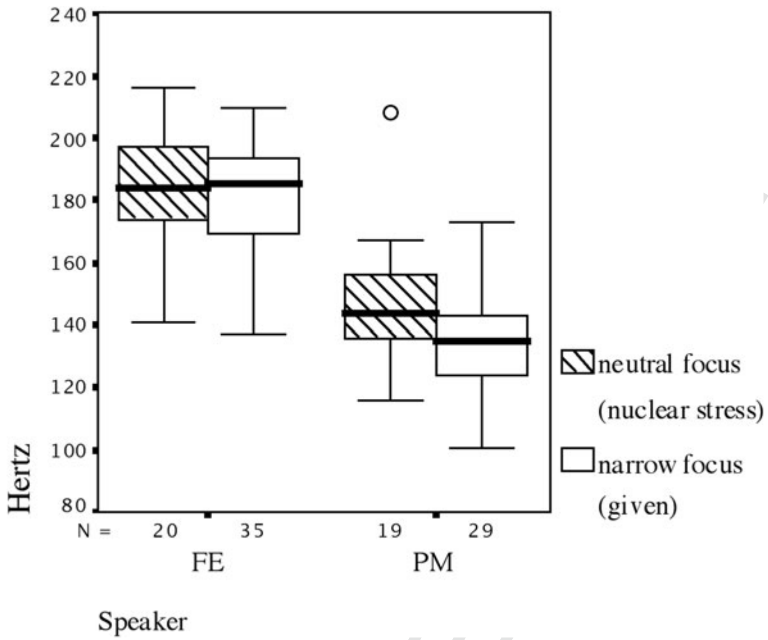


Fig. 11 Maximum right edge F0 by speaker and focus type

The present results did not support the null hypothesis that given material is deaccented in Thompson Salish. An independent samples t-test was significant for FE ($t=-7.017$, $df=53$, $p<.001$), allowing us to reject the null hypothesis that FE had lower F0 peaks on given material. For PM, a t-test was marginally significant ($t=-3.277$, $p=0.002$), suggesting that PM does not mark given material with perceptually salient lower F0 peaks either (Fig. 11).

F0 range (Fig. 12) was expected to be at least 3.6 semitones narrower on right edge given material, but this hypothesis was rejected for both speakers ($t=-6.434$, $df=53$, $p<0.001$, for FE; $t=-8.27$, $df=46$, $p<0.001$ for PM).

Differences in F0 peak timing were not markedly different between conditions. The mean differences (3% for FE, 13% for PM) are considerably less than the timing differences reported in previous studies (see Section 3.3), and so are unlikely to be of perceptual significance here. Because the data were not normally distributed (most F0 peaks occurred near the start of the vowel - Fig. 13) and did not have equal numbers across conditions, I did not perform t-tests for this variable.

The right edge: intensity (dB). Based on previous studies, deaccented vowels were expected to be 3 dB (or more) lower in their intensity peak. The results in the present study were inconclusive, since t-tests were not significant, so we cannot rule out that given material is marked through lower amplitude (Fig. 14). However, intensity is generally considered an unreliable cue since, from the listener's perspective, it is easily affected when, for example, a speaker turns her head.

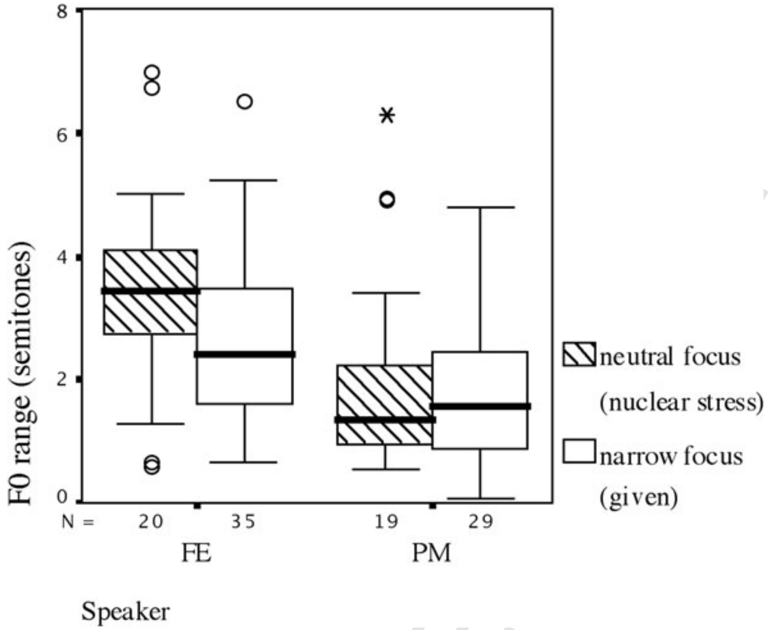


Fig. 12 Right edge F0 range by speaker and focus type

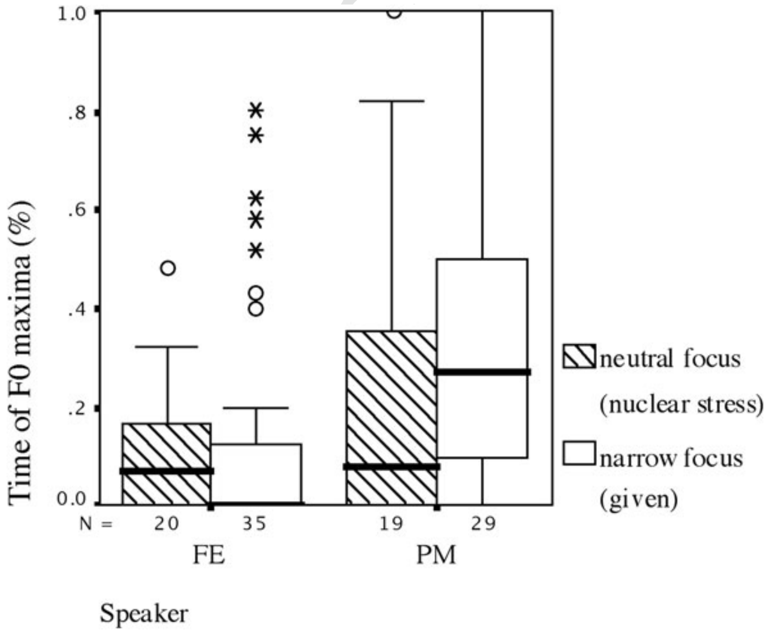


Fig. 13 Right edge time of F0 peak as a percentage of vowel duration

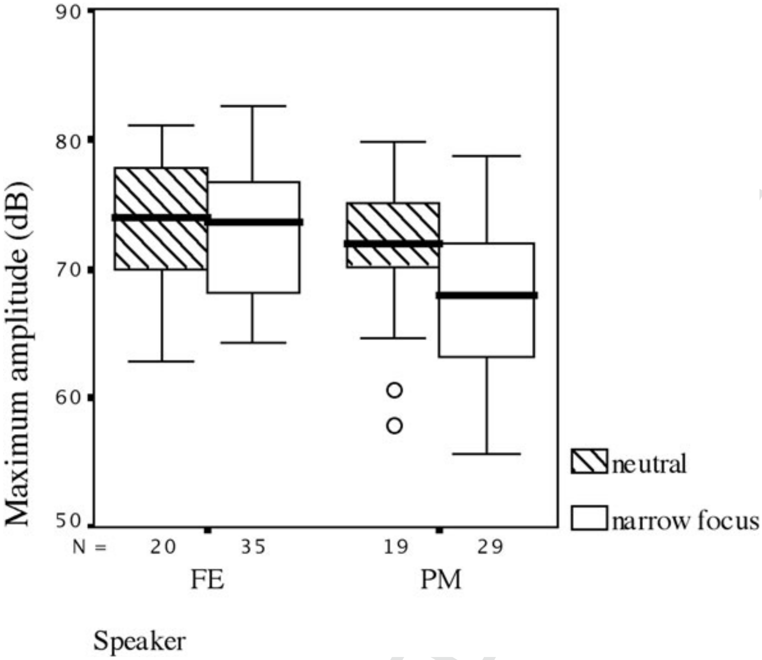


Fig. 14 Right edge peak amplitude (dB) by speaker and focus type

The right edge: duration (ms). We expected right-edge given vowels in the narrow focus condition to be at least 6% shorter (e.g. Sluijter and van Heuven 1996b). T-tests for this variable were not significant; we therefore cannot rule out that shorter duration is a cue for identifying given material. However, for PM, the durational difference between conditions was only 5.6%, unlikely to be perceptually salient. For FE, given vowels were on average 21% shorter, but considerable variability suggests that this cue is also not reliable (Fig. 15).

The right edge: summary. Given items in right edge cleft clauses were not marked by lower pitch height, lesser pitch excursion, or different pitch peak timing. Results for intensity and duration were not conclusive. Findings for each speaker are summarized in Tables 7 and 8.

4.6 Results: Declination

The declination from left to right: pitch. The difference in declination between focus types was not significant for either speaker, and in fact was less than the –2.2 semitones found in Eady and Cooper (1986: 407, calculated from Table III). For FE, this difference was –0.29 semitones, and for PM it was –1.27 semitones. Figures 16 and 17 show a graphic representation of each speaker's

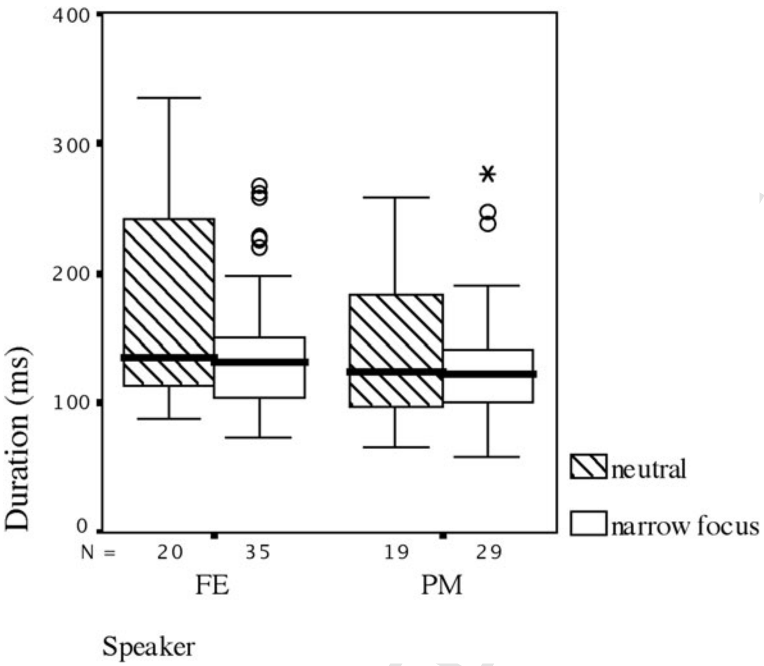


Fig. 15 Right edge stressed vowel duration (ms) by speaker and focus type

Table 7 FE rightmost stress: summary of acoustic cues, and t-test results

Measure	Focus	Mean (SD)	*n*	Null Hypoth.	*t*	*p*	*df*
–	–	–	–	–	–	–	–
Max F0 (Hz)	Neutral	183.4 (18.4)	20	$\mu_1 - \mu_2 \geq 3.5$	−7.017	***<0.001	53
	Narrow	180.2 (19.1)	35	semitones	–	–	–
–	–	–	–	–	–	–	–
F0 excursion	Neutral	3.50 (1.65)	20	$\mu_1 - \mu_2 \geq 3.6$	−6.434	***<0.001	53
(semitones)	Narrow	2.71 (1.49)	35	semitones	–	–	–
–	–	–	–	–	–	–	–
Time of	Neutral	0.11 (0.13)	20	$\mu_1 \neq \mu_2$	–	–	–
F0 peak (%)	Narrow	0.14 (0.24)	35	–	–	–	–
–	–	–	–	–	–	–	–
Maximum	Neutral	73.7 (4.83)	20	$\mu_1 - \mu_2 \geq 3$ db	0.855	>0.3	53
intens. (dB)	Narrow	72.5 (5.10)	35	–	–	–	–
–	–	–	–	–	–	–	–
V duration	Neutral	176.3 (78.2)	20	$\mu_1 - \mu_2 \geq .06{*}\mu_2$	1.22	>0.2	53
(ms)	Narrow	145.2 (57.4)	35	-	–	–	–

mean pitch contour (in Hertz), by focus type (compare the anticipated contour in Fig. 5). The whiskers indicate error bars of one standard deviation in either direction; error bars for neutral focus are dashed, while error bars for narrow focus are solid.

Table 8 PM rightmost stress: summary of acoustic cues, and t-test results

Measure	Focus	Mean (SD)	*n*	Null Hypoth.	*t*	*p*	*df*
–	–	–	–	–	–	–	–
Max F0 (Hz)	Neutral	147.4 (20.4)	19	$\mu_1 - \mu_2 \geq 3.5$	−3.277	**0.002	46
	Narrow	134.7 (16.4)	29	semitones	–	–	–
–	–	–	–	–	–	–	–
F0 excursion	Neutral	2.01 (1.66)	19	$\mu_1 - \mu_2 \geq 3.6$	−8.27	***<0.001	46
(semitones)	Narrow	1.75 (1.17)	29	semitones	–	–	–
–	–	–	–	–	–	–	–
Time of	Neutral	0.24 (0.23)	19	$\mu_1 \neq \mu_2$	–	–	–
F0 peak (%)	Narrow	0.37 (0.35)	29	–	–	–	–
–	–	–	–	–	–	–	–
Maximum	Neutral	71.6 (5.65)	19	$\mu_1 - \mu_2 \geq 3$ db	0.290	>0.2	46
intens. (dB)	Narrow	68.1 (5.79)	29	-	–	–	–
–	–	–	–	–	–	–	–
V duration	Neutral	141.8 (56.2)	19	$\mu_1 - \mu_2 \geq$	−0.035	>0.2	46
(ms)	Narrow	134.3 (57.3)	29	$0.06{*}\mu_2$	–	–	–

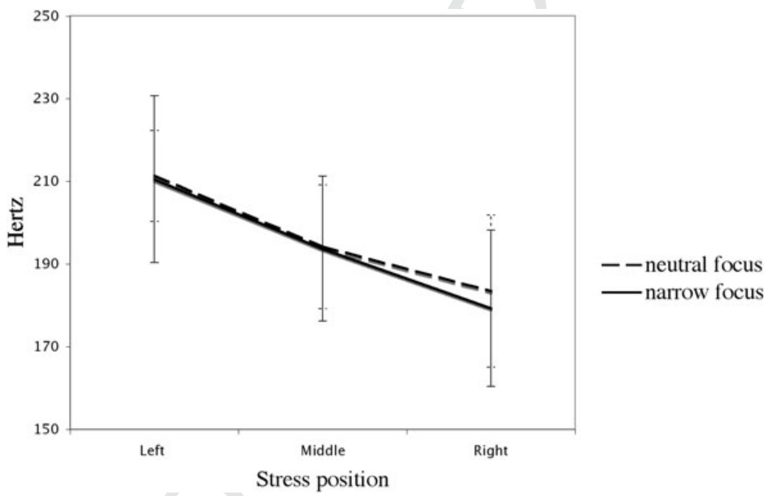


Fig. 16 Pitch contour across F0 peaks for FE, by focus type

A second F0 measurement to be checked was the difference in F0 range within each utterance between left and right stresses. For both speakers, the difference in means between the two conditions was non-significant ($t = 1.584$, $df = 52$, $p > 0.1$, for FE; $t = -1.583$, $df = 49$, $p > 0.1$, for PM).

The declination from left to right: intensity (dB). Means tended in the expected direction (FE had 3.06 dB and PM 2.59 dB greater amplitude declination, on average, in narrow focus utterances). However, even when the data from both speakers was combined, these differences were not significant ($t = 2.971$,

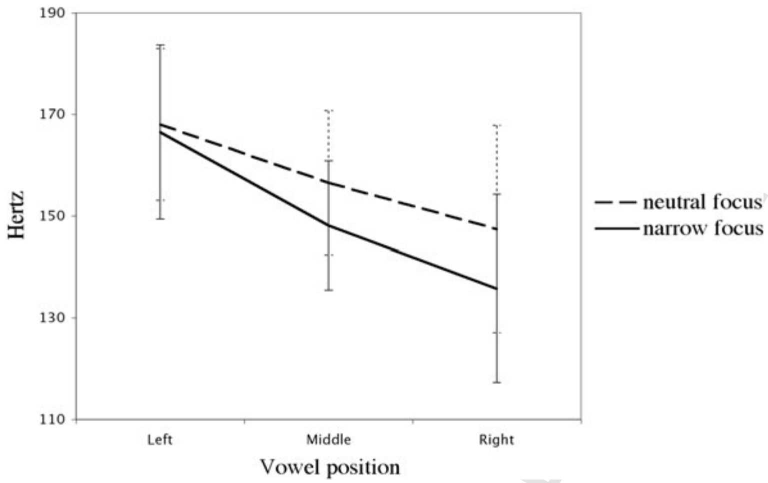


Fig. 17 Pitch contour across F0 peaks for PM, by focus type

Table 9 FE declination effects: summary of acoustic cues, and t-test results

Declination Measure	Focus	Mean (SD)	*n*	Null Hypoth.	*t*	*p*	*df*
–	–	–	–	–	–	–	–
ΔMax F0	Neutral	2.51 (1.93)	20	$\mu_1-\mu_2 = 0$	−0.582	>0.5	52
(semitones)	Narrow	2.80 (1.59)	34	–	–	–	–
–	–	–	–	–	–	–	–
ΔF0 range	Neutral	−0.56 (2.32)	20	$\mu_1-\mu_2 = 0$	−1.584	>0.1	52
(semitones)	Narrow	0.44 (2.17)	34	–	–	–	–
–	–	–	–	–	–	–	–
ΔMaximum	Neutral	−0.24 (4.18)	20	$\mu_1-\mu_2 = 0$	−2.971	**0.004	103
amplit. (db)	Narrow	3.06 (5.19)	34	–	–	–	–

Table 10 PM declination effects: summary of acoustic cues, and t-test results

Declination Measure	Focus	Mean (SD)	*n*	Null Hypoth.	*t*	*p*	*df*
–	–	–	–	–	–	–	–
ΔMax F0	Neutral	2.34 (1.86)	19	$\mu_1-\mu_2 = 0$	−1.906	>0.05	49
(semitones)	Narrow	3.61 (2.50)	32	–	–	–	–
–	–	–	–	–	–	–	–
ΔF0 range	Neutral	−0.50 (2.19)	19	$\mu_1-\mu_2 = 0$	−1.772	>0.05	49
(semitones)	Narrow	0.36 (1.66)	32	–	–	–	–
–	–	–	–	–	–	–	–
ΔMaximum	Neutral	1.25 (5.86)	19	$\mu_1-\mu_2 = 0$	−2.971	**0.004	103
amplit. (db)	Narrow	3.84 (4.50)	32	–	–	–	–

$df = 103$, $p = 0.004$), suggesting that narrow focus utterances are not reliably marked by greater amplitude declination.

The declination from left to right: summary. The analysis failed to uncover any significant cues to focus type (neutral or narrow) in the declination contours of the utterances Tables 9 and 10. This is consistent with the results in Sections 4.4 and 4.5.

4.7 Discussion

The most notable and robust finding in the present study is the complete absence of pitch cues in the marking of both narrowly focused items at the left edge, and given material at the right. Neither F0 peak nor F0 excursion were employed to mark information structure; this null hypothesis was rejected in all cases for both speakers (excepting one marginally significant result for PM). Since results were statistically significant, the absence of focus-induced F0 prominence marking is not due to noise in the data. The timing of F0 peaks was also not affected by the status of focus or givenness. Examination of the declination lines in wide focus utterances and narrow focus utterances also failed to detect the sort of dramatic pitch drops after the narrow focus constituent that is characteristic of languages like English or Hungarian. Absence of F0 cues appears to be typologically unusual for stress-accent languages like Nłeʔkepmxcin, but similar findings have been reported for the Niger-Congo language of Wolof (Rialland and Robert 2001), and the Papua New Guinean language of Kuot (Lindstrom and Remijsen 2005). Both are stress languages, yet fail to use pitch accents to mark information structure. The implication, also suggested by these authors, is that the role of phrasal stress in cueing focus marking has been overestimated by the study of focus in Indo-European languages.

The results for amplitude and duration in this study were inconclusive, so we cannot rule out that these phonetic cues are used to mark information structure – although, as an anonymous reviewer remarks, this too would be unusual for a stress language (but for a possible role of these phonetic cues in marking second occurrence focus in Germanic, see Beaver et al. 2007; Féry and Ishihara 2009). However, the failure to detect significant differences between focus types suggests that these factors are not used to cue focus and givenness in Thompson Salish. Most importantly, the lack of pitch cues for distinguishing non-given from given information indicates that pitch accents are not employed to mark information structure.

In terms of the correspondence of stress and focus, we then come to the following conclusion:

(12) Stress-Focus is not operative in Nłeʔkepmxcin:
Narrowly focused constituents do not attract greater prosodic prominence.

When it comes to the deaccenting of given information, we have found that given information is not deaccented in Nłeʔkepmxcin. This hypothesis would place Nłeʔkepmxcin together with other languages that exhibit a lack of deaccenting of old information (see Ladd 1996 for an overview).

(13) Destress-Given is not operative in Nłeʔkepmxcin:
Given information does not receive lesser prosodic prominence.

A weaker interpretation of the results would be that, if Nłeʔkepmxcin speakers do mark information structure by regulating acoustic prominence, they employ much more subtle phonetic cues than in languages previously studied. However, assuming similar perceptual abilities to speakers of other stress languages, I adopt the stronger interpretation of the results here.

5 Further Implications

The results of the phonetic study in Section 4 suggest that neither Stress-Focus nor Destress-Given are operative in the Nłeʔkepmxcin grammar. On the other hand, as pointed out by an anonymous reviewer, the findings are consistent with focus marking in some tone languages. For example, in Chichewa (Bantu), prosodic phrasing rather than stress-prominence is employed to mark focus (Kanerva 1990; Truckenbrodt 1999; Downing 2003). Zerbian (2007) argues that in Northern Sotho (also Bantu), neither prosodic prominence nor prosodic phrasing play a role in marking focus. While these studies explore tone languages, the present findings are unique in documenting in this sort of phonetic detail a similar absence of stress-cues in what otherwise appears to be a bona-fide stress language (Thompson and Thompson 1992; Egesdal 1984). Whether prosodic phrasing plays a role is a question beyond the scope of the present paper (but see Koch 2008a for some discussion).

If acoustic prominence is not relevant for marking focus, then we might expect this to be relevant in other areas of the focus marking system. For example, Beaver and Clark (2008) argue that focus sensitive operators that conventionally associate with focus (see Rooth 1985), such as *only, even*, and *also*, must associate with a stressed focus, since this is how focus is marked in English.

In Nłeʔkepmxcin, the focus associated with these operators does not appear to attract additional acoustic prominence. More spectacularly, the focus can simply be the phonologically null 3rd person pronoun *pro*. This is possible in simple clefts of the type seen in (4) and (5): in (14), the head of the cleft, the focus, is simply null *pro*. Where the English translation requires focal accent on the pronoun *SHE*, the Thompson Salish example simply has no overt exponent of the focus at all (but see Heim 1992, Krifka 1998, Rullmann 2003, for some cases of null foci in English).

(14) óo, c̓é=m̓=n̓=ekʷu [*pro*]$_{\text{FOC}}$ Ø=ʔex-st-émus cʔéył.
oh, CLEFT=EMPH=Q=EVID [3SG]$_{\text{FOC}}$ COMP=IMPF-TR-SUBJ.GAPNOW
'Oh, is [SHE]$_{\text{FOC}}$ taking care of him now?'
(more literally: 'Oh, is it [*pro*]$_{\text{FOC}}$ that is taking care of him now?')

In (15), the focus associated with the focus sensitive exclusive operator *ƛ̓uʔ* 'only' is null *pro*. This is not possible in English. However, in Nłeʔkepmxcin, where prosodic prominence is not relevant, such a construction is possible. On the other hand, note that a syntactic focus marking (a DP-cleft) *is* obligatorily employed to mark the focus associated with 'only' in (15). While 'only' associates with a conventionally marked focus in both English and Salish, the form of this marking is different: prosodically prominent in English, but syntactically marked in Salish (see Koch and Zimmermann to appear on focus sensitive operators more generally, Davis 2007 on Sƛ̓áƛ̓imcets Salish).

(15) cúkʷ=ƛ̓uʔ=xeʔ [*pro*]$_{\text{FOC}}$ e=rít-Ø-Ø-ne.
CLEFT$_{\text{ONLY}}$=ƛ̓uʔ=DEM [3SG]$_{\text{FOC}}$ COMP=read-TR-3O-1SG.S
'[It]$_{\text{FOC}}$ is the only one I am reading.'
(more literally: 'It is only [*pro*]$_{\text{FOC}}$ that I am reading.')

Since speakers are also fluently bilingual in English, a language in which information structure *is* marked through acoustic prominence, the results suggest that intonational properties and focus/givenness marking are not readily transferred from one language to another.

More broadly, the results point to the importance of cross-linguistic research in the area of focus marking, in order to establish both the cognitive universals and cross-linguistic areas of variation in the marking of information structure. The undertaking of this research is especially critical given the endangerment of many of the world's lesser studied languages, and the inability to glean this knowledge from most, if not all, existing grammars. While previous information structure research has tended to concentrate on languages in the Indo-European realm, the study of stress languages from other language areas may therefore prove especially fruitful here.[7]

6 Conclusion

The acoustic phonetic study presented here indicates that the discourse categories of focus and givenness are not marked through acoustic prominence in Thompson River Salish. The absence of the use of pitch accents to mark information structure does not conform to the Stress-Focus Correspondence Principle, or constraints like DESTRESS-GIVEN, suggesting that neither of these is a universal principle, even in stress languages.

[7] Thanks to three anonymous reviewers for help in clarifying this section.

Acknowledgments I am indebted to language consultants Flora Ehrhardt and Patricia McKay, without whom this research would not be possible. Many thanks also to audiences at TIE3 in Lisbon, and comments from Henry Davis, Eric Vatikiotis-Bateson, Hotze Rullmann, Doug Pulleyblank, Daniel Büring, and Monique Charest. The final version profited considerably from the careful work of three anonymous reviewers, whom I thank for their thoughtful comments and suggestions. This work has been supported by Jacobs and Kinkade Research Grants from the Whatcom Museum Foundation (Bellingham, Washington); a Social Sciences and Humanities Research Council of Canada Research Fellowship; and two Research Fellowships from the Deutscher Akademischer Austausch Dienst. All errors are my own.

References

Astruc, Lluïsa, and Pilar Prieto. 2006. Stress and accent: Acoustic correlates of metrical prominence in A. Botinis (ed.) Catalan. In *Proceedings of Experimental Linguistics-2006*, 73–76.

Beaver, David, and Brady Clark. 2008. *Sense and Sensitivity: How Focus Determines Meaning*. Oxford: Blackwell.

Beaver, David, Brady Clark, Edward Flemming, T. Florian Jaeger, and Maria Wolters. 2007. When semantics meets phonetics: Acoustical studies of second-occurrence focus. *Language* 83(2): 245–276.

Benkirane, Thami. 1998. Intonation in Western Arabic (Morocco). In Daniel Hirst and Albert Di Cristo (eds.) *Intonation Patterns: A Survey of Twenty Languages*, 345–359. Cambridge: Cambridge UP.

Benner, Allison. 2006. The prosody of Senchothen. Paper given at The 41st International Conference on Salish and Neighbouring Languages, University of Victoria.

Boersma, Paul, and David Weenink. 2007. Praat: doing phonetics by computer (Version 4.5.17) Computer program. http://www.praat.org/. Retrieved March 21, 2007.

Botinis, Antonis. 1998. Intonation in Greek. In Daniel Hirst and Albert Di Cristo (eds.) *Intonation Patterns: A Survey of Twenty Languages*, 288–310. Cambridge: Cambridge UP.

Butcher, Andrew, and Jonathan Harrington. 2003. An instrumental analysis of focus and juncture in Warlpiri. In M. Sole, D. Recasens, and J. Romero (eds.) *Proceedings of the 15th International Congress of Phonetic Sciences*, Barcelona.

Chen, Yiya, and Carlos Gussenhoven. 2008. Emphasis and tonal implementation in Standard Chinese. *Journal of Phonetics* 36: 724–746.

Cooper, William, Stephen Eady, and Pamela Mueller. 1985. Acoustical aspects of contrastive stress in question-answer contexts. *Journal of the Acoustical Society of America* 77: 2142–2156.

Davis, Henry. 2007. Prosody-focus dissociation and its consequences: The case of Salish. Paper presented November 10, 2007, Nagoya, Japan.

Davis, Henry, and Lisa Matthewson. 2009. Issues in Salish Syntax and Semantics. *Language and Linguistics Compass* 3/4: 1097–1166.

Davis, Henry, Lisa Matthewson, and Scott Shank. 2004. On the presuppositionality of clefts in Samish and St̓át̓imcets. In D.B. Gerdts and L. Matthewson (eds.) *Studies in Salish Linguistics in Honor of M. Dale Kinkade*, 100–117. Missoula: University of Montana Working Papers in Linguistics 17.

Downing, Laura. 2003. Stress, tone and focus in Chichewa and Xhosa. In Rose-Juliet Anyanwu (ed.) *Stress and Tone: The African Experience*. *Franfurter afrikanisitische Blätter* 15, 59–81. Koln: Rüdiger Koppe Verlag.

Eady, Stephen, and William Cooper. 1986. Speech intonation and focus location in matched statements and questions. *Journal of the Acoustical Society of America* 80: 402–415.

Eady, Stephen, William Cooper, Gayle Klouda, Pamela Mueller, and Dan Lotts. 1986. Acoustical characteristics of sentential focus: Narrow vs. broad and single vs. dual focus environments. *Language and Speech* 29: 233–251.

Edwards, Mary. 1954. *Cree: An Intensive Language Course*. Prince Albert: Northern Canada Evangelical Mission.

Egesdal, Steven. 1984. *Stylized Characters' Speech in Thompson Salish Narrative*. Ph.D. dissertation, University of Hawaii.

Féry, Caroline, and Shinichiro Ishihara. 2009. The phonology of second occurrence focus. *Journal of Linguistics* 45: 285–313.

Féry, Caroline, and Vieri Samek-Lodovici. 2006. Focus projection and prosodic prominence in nested foci. *Language* 82(1): 131–150.

Fry, D.B. 1958. Experiments in the perception of stress. *Language and Speech* 1: 126–152.

Gårding, Eva. 1998. Intonation in Swedish. In Daniel Hirst and Albert Di Cristo (eds.) *Intonation Patterns: A Survey of Twenty Languages,* 112–130. Cambridge: Cambridge UP.

Gerdts, Donna. 1988. *Object and Absolutive in Halkomelem*. New York: Garland.

Grønnum, Nina. 1998. Intonation in Danish. In Daniel Hirst and Albert Di Cristo (eds.) *Intonation Patterns: A Survey of Twenty Languages,* 131–151. Cambridge: Cambridge UP.

Gumperz, John. 1982. *Discourse Strategies*. Cambridge: Cambridge UP.

Gussenhoven, Carlos. 2004. *The Phonology of Tone and Intonation*. Cambridge: Cambridge UP.

Halliday, Michael. 1967. Notes on transitivity and theme in English (part 2). *Journal of Linguistics* 3: 199–244.

Hartmann, Katharina. 2007. Focus and tone. In C. Féry, G. Fanselow and M. Krifka (eds.) *The Notions of Information Structure*. *Interdisciplinary Studies on Infomation Structure 6,* 221–235. Potsdam: Universitätsverlag Potsdam.

Hayes, Bruce, and Aditi Lahiri. 1991. Bengali intonational phonology. *Natural Language & Linguistic Theory* 9: 47–96.

Heim, Irene. 1992. Presupposition projection and the semantics of attitude verbs. *Journal of Semantics* 9: 183–221.

Jackendoff, Ray. 1972. *Semantic Interpretation in Generative Grammar*. Cambridge: MIT Press.

Kanerva, Jonni M. 1990. *Focus and Phrasing in Chichewa Phonology*. New York: Garland.

Keppel, Geoffrey, and Thomas Wickens. 2004. *Design and Analysis: A Researcher's Handbook*. Upper Saddle River, N.J.: Pearson Prentice Hall.

Kinkade, M. Dale. 1990. Sorting out third persons in Salish discourse. *International Journal of American Linguistics* 56: 341–360.

Kinkade, M. Dale. 1992. Salishan languages. In William Bright (ed.) *International Encyclopedia of Linguistics*, 359–362. New York: Oxford University Press.

Koch, Karsten. 2007. Focus projection in Nłeʔkepmxcin (Thompson River Salish). In Hannah Haynie and Charles Chang (eds.) *Proceedings of WCCFL 26*, 348–356. Somerville, MA: Cascadilla Proceedings Project. http://www.lingref.com/cpp/wccfl/26/paper1690.pdf

Koch, Karsten. 2008a. Intonation and Focus in Nłeʔkepmxcin (Thompson River Salish). Ph.D. dissertation, University of British Columbia. https://circle.ubc.ca/handle/2429/2848

Koch, Karsten. 2008b. Some issues in the structure and interpretation of clefts in Nłeʔkepmxcin (Thompson River Salish). In John Lyons (ed.) *Papers for the 43rd ICSNL*, Vancouver: UBC Working Papers in Linguistics.

Koch, Karsten, and Malte Zimmermann. To appear. Focus sensitive operators in Nłeʔkepmxcin (Thompson River Salish). In M. Prinzhorn, V. Schmitt and S. Zobel (eds.) *Proceedings of Sinn und Bedeutung 14*.

Krifka, Manfred. 1998. Additive particles under stress. In D. Strolovich and A. Lawson (eds.) *Proceedings from Semantics and Linguistic Theory VIII*, 111–128. Ithaca, NY: CLC Publications.

Kroeber, Paul. 1997. Relativization in Thompson Salish. *Anthropological Linguistics* 39(3): 376–422.

Kroeber, Paul. 1999. *The Salish Language Family: Reconstructing Syntax*. Lincoln: University of Nebraska Press.

Ladd, D. Robert. 1996. *Intonational Phonology*. Cambridge: Cambridge UP.

Lieberman, Philip. 1967. *Intonation, Perception, and Language*. Cambridge: MIT Press, Research Monograph No. 38.

Lindstrom, Eva, and Bert Remijsen. 2005. Aspects of the prosody of Kuot, a language where intonation ignores stress. *Linguistics* 43(4): 839–870.

Muehlbauer, Jeff. 2005. Pitch as accent in Nêhiyawêwin (Plains Cree). Paper presented at the 37th Algonquian Conference. Ottawa, Ontario.

Okobi, Anthony. 2006. *Acoustic Correlates of Word Stress in American English*. Ph.D. Dissertation, MIT.

Reinhart, Tanya. 2006. Interface Strategies. *Optimal and Costly Interpretations*. Cambridge: MIT Press Linguistic Inquiry Monographs.

Reinhart, Tanya. 1995. Interface strategies. *OTS Working Papers in Linguistics*. Utrecht: Utrecht University OTS.

Rialland, Annie, and Stéphane Robert. 2001. The intonational system of Wolof. *Linguistics* 39(5): 893–939.

Rochemont, Michael. 1986. *Focus in Generative Grammar*. Amsterdam: Benjamins.

Rooth, Mats. 1985. *Association with Focus*. Ph.D. Dissertation, UMass, Amherst.

Rooth, Mats. 1992. A Theory of Focus Interpretation. *Natural Language Semantics* 1: 75–116.

Rullmann, Hotze. 2003. Additive particles and polarity. *Journal of Semantics* 20: 329–401.

Schwarzschild, Roger. 1999. Givenness, AvoidF, and other constraints on the placement of accent. *Natural Language Semantics* 7: 141–177.

Selkirk, Elizabeth. 1995. Sentence prosody: Intonation, stress and phrasing. In J. Goldsmith (ed.) *The Handbook of Phonological Theory*, 550–569. Cambridge, MA: Blackwell.

Shih, Chilin. 1997. Declination in Mandarin. In A. Botinis, G. Kouroupetroglou, and G. Carayannis (eds.) Intonation: Theory, Models and Applications. *Proceedings of an ESCA Workshop*, 293–296. Athens: ESCA and University of Athens Department of Informatics.

Shue, Yen-Liang, Markus Iseli, Nanette Veilleux, and Abeer Alwan. 2007. Pitch Accent versus Lexical Stress: Quantifying Acoustic Measures Related to the Voice Source. In H. van Hamme and R. van Sonne (eds.) *Proceedings of Interspeech 2007*, 2625–2628.

Sluijter, Agaath, and Vincent van Heuven. 1996a. Acoustic correlates of linguistic stress and accent in Dutch and American English. *Proceedings of ICSL, 1996*, 630–633. Philadelphia: Applied Science and Engineering Laboratories, Alfred I duPont Institute.

Sluijter, Agaath, and Vincent van Heuven. 1996b. Spectral balance as an acoustic correlate of linguistic stress. *Journal of the Acoustical Society of America* 100(4): 2471–2485.

Strik, Helmer, and Loe Boves. 1995. Downtrend in F0 and Psb. *Journal of Phonetics* 23: 203–220.

Suomi, Kari, Juhani Toivanen, and Riikka Ylitalo. 2003. Durational and tonal correlates of accent in Finnish. *Journal of Phonetics* 31: 113–138.

Swerts, Marc, Eva Strangert, and Mattias Heldner. 1996. F0 declination in read-aloud and spontaneous speech. In *ICSLP-1996*, 1501–1504.

’t Hart, Johan. 1981. Differential sensitivity to pitch distance, particularly in speech. *Journal of the Acoustical Society of America* 69(3): 811–821.

’t Hart, Johan, René Collier, and Antonie Cohen. 1990. *A Perceptual Study of Intonation*. Cambridge: Cambridge UP.

Thompson, Laurence, and M. Terry Thompson. 1992. *The Thompson Language*. Missoula: University of Montana Occasional Papers in Linguistics 8.

Thompson, Laurence, and M. Terry Thompson. 1996. *Thompson River Salish dictionary*. Missoula: University of Montana Occasional Papers in Linguistics 12.

Truckenbrodt, Hubert. 1995. *Phonological Phrases: Their Relation to Syntax, Focus and Prominence*. Ph.D. dissertation, MIT.
Truckenbrodt, Hubert. 1999. On the relation between syntactic phrases and phonological phrases. *Linguistic Inquiry* 30: 219–255.
Turk, Alice, and Laurence White. 1999. Structural influences on accentual lengthening in English. *Journal of Phonetics* 27(2): 171–206.
van Kuppevelt, Jan. 1994. Topic and comment. In Ronald E. Asher (ed.) *The Encyclopedia of Language and Linguistics,* 4629–4633. Oxford: Pergamon Press.
Zerbian, Sabine. 2007. Investigating prosodic focus marking in Northern Sotho. In K. Hartmann, E. Aboh, and M. Zimmermann (eds.) *Focus strategies: Evidence from African Languages*, 55–79. Berlin: Mouton de Gruyter.

The Alignment of Accentual Peaks in the Expression of Focus in Korean

Kyunghee Kim

1 Introduction

In the phonological model of Korean intonation (see 1.1.1), intonation is defined by two tonally defined prosodic units, Accentual Phrase (AP) and Intonation Phrase (IP). AP is demarcated by the tonal pattern THLH (T = L or H) and IP by one of the nine boundary tones. The AP contours contain, as a rule, a peak which is assumed to be the phonetic reflex of the initial H tone associated with the second syllable of an AP. The AP peak is located either in the second or the third AP syllable. Wrong peak alignment may be perceived dialect coloured or slightly unnatural in certain contexts. However, most of the times, the alignment variation seems purely phonetic within the limited range of the second and the third syllables, and it is not yet known what conditions or affects the precise alignment.

Recent research has shown that various factors may influence the alignment of F0 peaks and valleys, such as, focus type, the location of the tonal event in a prosodic unit and the structure of a prosodic unit. Focus type is reported to affect the peak alignment of the nuclear fall in European Portuguese (Frota 2002). The declarative contour in European Portuguese is made up of an initial rise and a fall (HL) in the last stressed syllable of the intonational phrase with a high plateau in between. Frota shows that the beginning and the end of the fall (i.e., H and L) are aligned later in narrow focus than in broad focus. In broad focus, the fall starts in the syllable immediately preceding the stressed syllable and ends in the stressed syllable. On the other hand, in narrow focus the fall starts in the stressed syllable and ends in the following unstressed syllable. The peak alignment is also affected by the location of the fall in the intonational phrase. When an intonational phrase starts with the

K. Kim (✉)
IfL-Phonetics, University of Cologne, Cologne, Germany
e-mail: kimk@smail.uni-koeln.de

S. Frota et al. (eds.), *Prosodic Categories: Production, Perception and Comprehension*, Studies in Natural Language and Linguistic Theory 82,
DOI 10.1007/978-94-007-0137-3_7,

nuclear fall[1], the peak may also be aligned one syllable later than described above. That is, in the neutral fall, the peak is placed in the stressed syllable, not in the preceding syllable, and in the focus fall, it is located in the post-stressed syllable. The displacement of the peak has distinct effect on the alignment of the low end in the neutral and focal falls. In the neutral fall, the end of the fall is still aligned in the stressed syllable and the falling movement occurs solely in the stressed syllable. In the focal fall, it shifts together with the peak and occurs in the next unstressed syllable (i.e., the second unstressed syllable after the stressed syllable).

Focus type is also reported to affect the accent peak alignment in German (Baumann et al. 2007; Braun 2007). The nuclear accent peak is aligned later in contrastive accent than in non-contrastive accent; the peak is aligned later in contrastive focus than in broad focus or narrow focus (that is, broad <contrastive, narrow<contrastive). There was a strong tendency that the peak latency increases in the order of broad < narrow < contrastive, indicating that, possibly, different focus structures are reflected in the different degrees of peak latency. However, it was not found statistically significant (Baumann et al. 2007). Similar reports are made on German pre-nuclear accent that contrastive context causes later peak alignment (Braun 2006, 2007).

Whereas narrow (and contrastive) focus causes later peak alignment in European Portuguese and German, it is reported that the rising accent in European Spanish shows earlier peak alignment in narrow focus than in broad focus. Face (2001) shows that the peak in the rising accent occurs in the stressed syllable in narrow focus, whereas it occurs in the syllable following the stressed syllable in broad focus. He also notes that the accent location affects the speaker's decision on the narrow focus strategy. The earlier peak alignment is more frequently used for the focus distinction in the utterance initial accent. In non-initial accents, speakers tend to opt for raising the peak height rather than varying the peak alignment.

On the other hand, the alignment of the pitch turning point is affected by the structure of a prosodic unit. In French intonation, the AP tonal pattern is defined as LHiLH (Jun and Fougeron 2000). The initial L has been assumed to be associated with the initial syllable of the first content word in an AP. However, Welby (2003) makes an observation that the beginning of the AP initial LH rise may be aligned earlier. The initial low pitch elbow (created by the LHi) occurs at the boundary between the function word and the content word when the word is preceded by a function word.

In this study, we hypothesised that the variation in the alignment of the accentual peak in Korean is systematic and linguistically conditioned, and investigated if the peak alignment is affected by different linguistic factors. In the first

---

[1]Frota (2002) provides the following examples of the neutral and the focal falls in the intonational phrase initial position. Stressed syllable is in capitals, narrowly focused item in bold and the expected intonational phrasing is indicated with square brackets and 'I'.

Neutral : [As angoLAnas]I [ofereceram especiarias aos jornalistas]I
'The Angolans gave spices to the journalists'
Focal : [As angoLAnas ofereceram especiarias aos jornalistas]I

production experiment, we tested the three factors that have been reported to affect the alignment of peaks and valleys, namely, focus type, the AP location (represented with the presence/absence of the preceding AP) and the AP structure (represented with the presence/absence of the second phonological word). In the second experiment, based on the outcome of the first experiment, we have examined the possible effect of the morpheme boundary and the presence of semantic content in the second morpheme on the peak alignment.

In the next sections, we will first give a brief description of Korean intonation proposed by Jun (1996, 2000) and of the stress and its domain (i.e., phonological word) in Lee (1990) and discuss how stress and phonological word can affect the AP peak alignment. This is followed by the detailed report of the two production experiments (Section 2) where the variation in the AP peak alignment was investigated by measuring the distance from the beginning of the second AP syllable to the accentual peak. In Section 3, the findings of the experiments and their theoretical implications are discussed.

1.1 Intonational Structure of Seoul Korean

In this section, we will first look at the structure of Korean intonation proposed by Jun (1996) and Korean ToBI (Jun 2000) as well as its labelling conventions of AP contours in the phonetic tone tier. These works provide the basis of the prosodic analysis in the following production experiments. Moreover, in this paper, we will adopt the labelling conventions in K-ToBI phonetic tone tier to illustrate the AP contours, as they provide faithful representations of the contours without losing details, such as, the syllable location of peaks and valleys (cf. rise-fall-rise). We shall, then, examine the stress rule and the domain of the stress assignment proposed in Lee (1990) and consider how stress might affect the realisation of AP tones.

1.1.1 Jun (1996) and Korean ToBI (Jun 2000)

Jun's intonation model assumes the intonational phonology proposed in Pierrehumbert (1980), Beckman and Pierrehumbert (1986) and Pierrehumbert and Beckman (1988). The prosodic units in the model are hierarchically structured following the Strict Layer Hypothesis (Selkirk 1984) and higher level constituents are exhaustively parsed into one or more of the immediately subordinating components. Two tonally defined prosodic units, the Intonational Phrase (IP) and the Accentual Phrase (AP), are assumed. The IP is characterised and delimited by the obligatory IP boundary tone on the last syllable of an IP. So far, nine different tonal complexes are identified; L%, H%, LH%, HL%, LHL%, HLH%, HLHL%, LHLH%, LHLHL%. The syllable associated with the IP boundary tone is lengthened about 1.8 times (final lengthening) when compared to the phrase initial syllable (Lee and Seong 1996).

The AP is demarcated by its tonal pattern THLH, where T represents either H or L according to the presence or absence of the segmental feature [stiff vocal cords] in the onset of the initial syllable. The presence of the feature [stiff vocal cords] places fortis, /p', t',k', ts'/, strongly aspirated obstruents, /p^h, t^h, k^h, ts^h/ and fricatives[2], /s^h, s, h/ under one category and T represents H when the onset is [+stiff vocal cords]. On the other hand, the absence of the feature groups the lenis consonants, /p, t, k, ts/, and sonorants together and T represents L when the onset (of the initial syllable) lacks [stiff vocal cords].

In Jun (1996), she assumes a series of rules which map and allocate the AP tones, i.e., THLH, to individual TBUs (Tone Bearing Units), so that every TBU in an AP is associated with at least one tone. In Korean ToBI, however, it appears only the two initial and the two final syllables of an AP are assumed to be associated with tones. Fig. 1, taken from Jun (2000), shows the prosodic hierarchy and the intonation structure in Seoul Korean. The hierarchy shows that the highest level prosodic constituent, that is, an IP, consists of one or more APs, an AP of one or more phonological words and a phonological word of one or more syllables. The TH of the THLH is represented to be associated with the two initial syllables and the LH with the two final syllables of an AP showing no tonal specification for the other syllables. As represented in Fig. 1, the tones may be assigned to the syllables that belong to different phonological words within the AP.

When an AP is Intonation Phrase final, the tonal specification of the final syllable is 'pre-empted' by the mandatory IP boundary tone. In Fig. 1, this is represented by the lines connecting the final syllable to both the AP final H and %, an IP boundary tone, which is also linked to the IP on the top of the prosodic tree. If an AP containing four syllables is IP final and displays, for example, the tonal pattern of LHLH, then it may be analysed as AP tones, LHL, followed by the IP boundary tone, H%.

Unlike other ToBI systems, Korean ToBI has two different tone tiers, a phonological tone tier and a phonetic tone tier. This is due to the fact that in K-ToBI and Jun (1996), various AP contours (see Fig. 2) are assumed to be derived from the identical underlying AP tonal pattern of THLH and yet there are contour shapes which are not predictable from the underlying tonal pattern[3]. For that reason, on the phonetic tone tier the AP contours are precisely described by specifying F0 levels and turning points at up to four different locations in an AP (see Fig. 2): the F0 valley or peak in the AP initial syllable is labelled with L(ow) or H(igh); the high turning point on the second (or the

---

[2] The (voiceless) alveolar fricatives are traditionally classified as lenis and fortis and transcribed as /s/ and /s'/, respectively (e.g., Huh 1985). In Jun's work, however, the lenis is transcribed as /s^h/ and, therefore, classified as [+stiff vocal cords]. For further details, the reader is referred to Jun (1996).

[3] Expanding the K-ToBI tone tier to add a phonetic tone tier, Jun (2000) explains that the decision was made 'in order to describe surface tonal patterns which are not predictable from the underlying tones' and 'to investigate if there is any meaning difference among [the varying contours]'

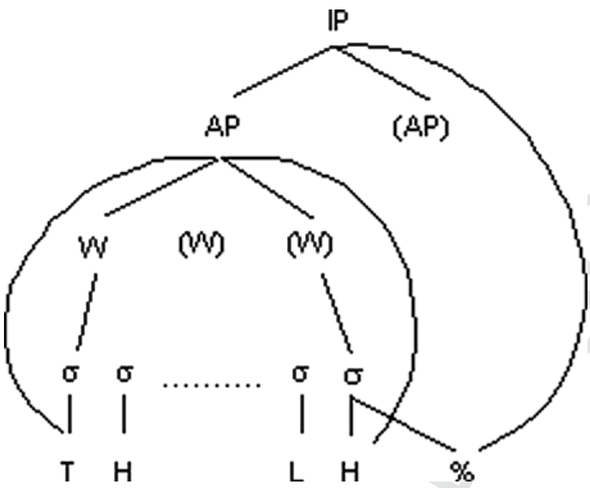


Fig. 1 Intonational structure of Seoul Korean (taken from Jun 2000)
IP: Intonation Phrase, AP: Accentual Phrase, w: phonological word, s: syllable, T= H, when the syllable initial segment is aspirated/tense, otherwise, T= L, %: Intonation phrase boundary tone

third) syllable with +H; the low turning point on the AP penultimate syllable with L+; and the F0 target in the final syllable with La or Ha. In this paper, we adopt the labelling conventions of K-ToBI phonetic tone tier to illustrate the AP contours for their faithful depiction of the contours.

It should be noted that Jun stipulates that La is to be used to indicate the F0 valley in the AP final syllable. Nonetheless, in this paper, we took the liberty of employing La to describe the low falling F0 in the final syllable. The target APs in Experiment 1 and 2 display rising-falling contours. The falling portion of the

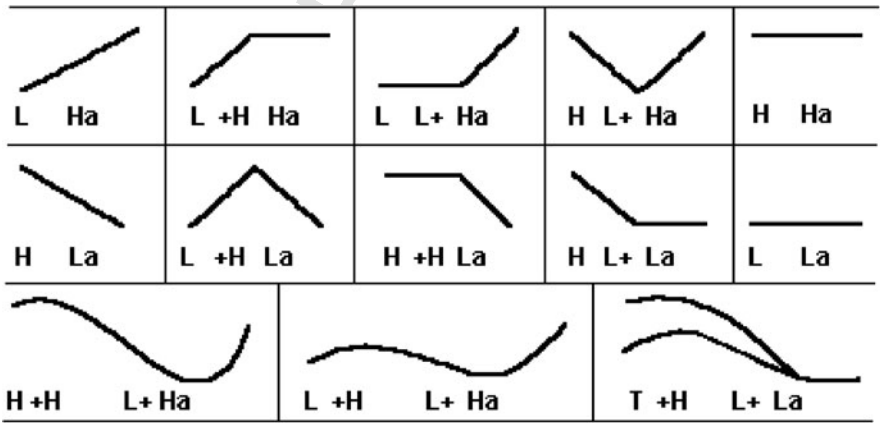


Fig. 2 Schematic representations of the labelled AP contours on the phonetic tone tier (taken from Jun 2000)

contour has no F0 target in the final syllable of the AP, but simply continues to fall to the initial syllable of the following AP, where the F0 valley is located. If we abide by the original convention, the AP contour would be labelled as L+H, as there is no F0 target in the final syllable. This may give the wrong impression that it is a rising contour when it actually is a rising-falling contour. For that reason, the rising-falling contour is represented as L+HLa despite the lack of the low F0 turning point in the final syllable.

1.1.2 Stress in Lee(1990)'s Korean Intonation Model and Phonological Words

Lee's model is built from the theoretical perspective of the British school, represented by O'Connor and Arnold (1973) and Crystal (1969). In the model, even though Korean lacks lexical stress, Lee postulates abstract word-level 'stress'. His stress rule in (1) assigns 'stress' primarily on the first syllable of a morpheme excluding 'clitics'. Clitics, which includes morphemes such as endings, (most) prefixes and suffixes, postpositions (i.e., particles), bound nouns and bound predicates, do not have a stress of their own and form a phonological unit together with the preceding (or, in the case of prefixes, the following) morpheme(s) to which the stress rule is applied. The rule basically assigns stress to the initial syllable of a phonological word which is the prosodic unit just below AP in Jun's model (see 1.1.1). According to Lee, this 'stress' is abstract and phonological in the sense that it is a 'pre-condition for potential accent' (Lee 1990: 19) and has no phonetic manifestation, unless it is 'accented' and receives pitch prominence by starting an AP. In other words, a stressed syllable is the initial syllable of a phonological word and the location where an AP may start in broad/neutral focus.

(1) The Korean Stress Rule (Lee 1990: 50)[4]

1) Two syllable morphemes:
Stress falls on the first syllable

2) Three or more syllable morphemes:
If the first syllable is heavy, stress falls on that syllable. Otherwise, either on the first or on the second syllable with no important linguistic difference implied.

[4]It should be noted that the second part of the stress rule allows accent (i.e., pitch prominence) to fall on the second syllable of a morpheme. This is possible, as a 'rhythm unit', AP level prosodic unit in Lee's model, may contain anacrusis, unstressed syllable(s) preceding a stress. A rhythm unit is defined as optional anacrusis, an obligatory stressed/accented syllable and the following (optional) unstressed syllables. It is not clear what Lee indicates with 'important linguistic difference' in his stress rule. However, comparing the two possible stress patterns of /tsa.doŋ.tɕ^{h}a/ 'automobile, car' and /kɛ.gu.ɾi/ 'frog', he acknowledges that placing stress (and accent) on the second syllable makes them sound emphatic (1990: 47-48). This suggests that phonological stress should fall exclusively on the first syllable of a morpheme regardless of the syllable count and, therefore, that an AP should actually start at the initial syllable of a phonological word.

In Lee's model, stressed syllables, i.e., phonological word initial syllables, are metrically strong syllables placed low in the prosodic hierarchy. Lee explains that accentual phrasing (accent placement in Lee's terms) in broad focus sentences is governed by the prosodic structure of the sentence as well as the factors, such as, the scope of focus, speech rate and style. The prosodic structure represents 'prosodic constituency and prosodic strength relations (different degrees of stress)' (Lee 1990: 72) and it is hierarchical in that the node branching from higher in the structure tree is more likely to be accented, i.e., phrased, than the node lower in the tree. He claims that, for example, /ma.ɾi. maː.nɯn. sa.ɾam/ '(A) talkative person' has a prosodic structure as illustrated below (see Figure 3). It has three phonological words, /ma.ɾ -i/ 'language, word'-subject particle, /maː.nɯn/ 'many' and /sa.ɾam/ 'person'. /ma.ɾ -i/, a strong node branching directly from the top node, is most likely to be accented when uttered, followed by /maː.nɯn/. /sa.ɾam/ is least likely to be accented, as it is a weak node branching from another weak node.

Lee explains that stress is a phonological entity without phonetic realisation and is assumed in order to predict neutral accentual phrasing. However, Cho and Keating (2001) suggests that stressed syllables, i.e., phonological word initial syllables, may be characterised with stronger segments than unstressed syllables. They investigated the strengthening of the prosodic domain initial alveolar consonants by examining a number of articulatory and acoustic parameters. Their results indicate that the onsets of the phonological word initial syllables are produced with greater peak linguopalatal contact (in /n/), longer stop seal duration (in /n/ and /t/) and longer VOT (in /tʰ/[5]), than those of the syllables located (phonological) word mediallly. Considering that the phonetic correlates of stress can hardly be defined even in languages like English, in which stress is clearly perceived, Cho and Keating's work suggests that Lee's stress has phonetic realisation on the segmental level and, possibly, also on the intonation level.

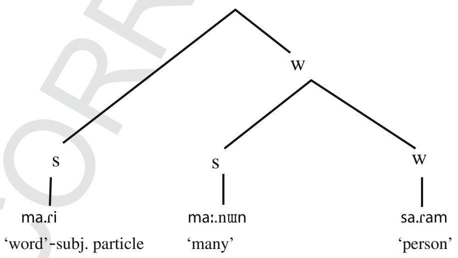


Fig. 3 Prosodic structure of / ma.ɾi. maː.nɯn. sa.ɾam / '(A) talkative person' (taken from Lee 1990: 72)

[5]Note that measuring VOT (voice onset time) does not apply to /n/ and /t/ among the tested alveolar consonants. Particularly, /t/ is voiced when located AP medially and it is voiced both in the phonological word initial and medial syllables. On the other hand, /t'/ does not display the identical characteris-tics as / tʰ/.

1.2 Phonological Words and the Accentual Peak Alignment

An AP usually contains one or two phonological words with five syllables or less. In an AP with two phonological words, the first word is typically two or three syllables long. Note that the accentual peak (i.e., the initial H peak of the THLH pattern) falls in the second or the third syllable of an AP, which suggests that the presence of the second stress (on the initial syllable of the second phonological word) is likely to affect the alignment of the accentual peak. Even though stressed syllable is assumed metrically strong, the peak does not seem to fall on the second stressed syllable. Rather, it seems that the AP peak is restrained to be aligned immediately before the second stressed syllable. Consider [[mi.ra]$_{PW}$ [ʌn.ni]$_{PW}$]$_{AP}$ and [[mi.ra.n -i]$_{PW}$ [ʌn.ni]$_{PW}$]$_{AP}$, for instance. The two APs are very similar segmentally and structurally. They both consist of two phonological words; a girl's name /mi.ra/, or / mi.ran/, and the following /ʌn.ni/ 'elder sister' meaning 'elder sister Mira' and 'elder sister Miran', respectively. The difference lies in the number of syllables in the first phonological word. [mi.ra.n -i]$_{PW}$ in the second AP contains three syllables, as the suffix /-i/, which helps the pronunciation (Yonsei Institute of Language and Information Studies 1998), causes the resyllabification of /ran / in /mi.ran/. Note that the accentual peak occurs at the end of the first phonological word in both the APs (see the illustration below). The peak is aligned in the second syllable in [[mi.ra]$_{PW}$ [ʌn.ni]$_{PW}$]$_{AP}$ where the first word contains two syllables. On the other hand, it is aligned in the third syllable in [[mi.ra.n -i]$_{PW}$ [ʌn.ni]$_{PW}$]$_{AP}$ where the first word contains three syllables. Also consider the one-word AP [[mi.ra.-han.t^{h}e]$_{PW}$]$_{AP}$ 'Mira'$_{\text{dative case marker}}$, which has the peak in the second syllable.

(cf.)

[[mi.ra]$_{PW}$ [ʌn.ni]$_{PW}$]$_{AP}$ [[mi.ra. -i]$_{PW}$ [ʌn.ni]$_{PW}$]$_{AP}$ [[mi.ra. -han.t^{h}e]$_{PW}$]$_{AP}$

The alignment of the AP peak in Korean is similar to that of the L in the AP initial LH rise (LHi) in French (Welby 2003) in that they are aligned near the boundary of a word level unit and mark the edge of the unit; the accentual peak in Korean is aligned at the end of the first phonological word in APs with 'phonological word-phonological word' structure and the low F0 turning point in French is aligned at the beginning of the content word in APs with 'function word-content word'. The 'function word-content word' structure in French APs is comparable to the 'phonological word-phonological word' in Korean APs in the sense that the APs contain a sequence of word level units. The structure is also comparable to a Korean one-word AP with a 'content morpheme-functional morpheme' structure in the sense that one element has semantic content and the other is functional. The French example suggests that the accentual peak alignment in Korean is likely to be affected by the presence of the second phonological word or the location of the initial morpheme boundary.

We assume that prosody is more likely to be affected by the prosodic structure of an AP, that is, by the presence of the second phonological word, than the morphological structure of a prosodic unit. We start the production experiment by testing the assumption. The possible effect of morphological structure on the peak alignment is left to be dealt with in the second experiment. We assume that the accentual peak is as default aligned with the second AP syllable which the accentual H tone is associated with. It is also assumed that the peak is aligned at the end of the first word, when an AP contains two phonological words (or more). We assume that the stress in the second phonological word attracts the peak to be aligned in the stressed syllable but the alignment is restrained by the phonological word boundary. This assumption on the peak alignment will be tested together with the assumption that the peak alignment is also affected by sentence length (i.e. the presence/absence of the preceding AP) and focus type.

2 Production Experiments

Under the assumption that the variation in the accentual peak alignment is systematic and linguistically conditioned, we attempted to seek the answer to the question, what conditions the peak alignment in Korean? As the accentual peak is assumed to be the phonetic reflex of the H tone associated with the second syllable of an AP, we measured the distance from the beginning of the second syllable to the peak and examined if the AP peak alignment is affected by the factors; the AP structure, which is varied with the number of phonological words in an AP; the AP location in an utterance, varied by the presence or the absence of the preceding AP; and focus type. Following the results of the investigation, in Experiment 2, we additionally examined the influence of the morpheme boundary location and the presence of semantic content in the second morpheme on the peak alignment.

It should be noted that we restricted our investigation of the peak alignment to one type of AP contour, L+HLa (see Fig. 2). Such measure was taken in order to prevent the variation that is induced by the different contour shapes. In Jun's intonation model (see 1.1.1), which provides the theoretical basis for this work, it is hypothesised that all the varying pitch contours shown in Fig. 2 are mere phonetic variants of the identical AP tonal pattern THLH. However, the alignment of peaks and valleys may vary among the variants due to the realisation of different tonal targets. For instance, all else being equal, the accentual peak (+H) alignment in L+HL+Ha may be earlier than in L+H Ha due to the realisation of the second L tone. Unlike in L+HHa where only three of the four AP tones are realised, in L+HL+Ha all four are realised. The presence of the pitch valley (L+) may force the initial peak (+H) to be earlier than in L+HHa, as the realisation of the second L tone requires more room.

Among the varying contour shapes, we opted for L+HLa (see 1.1.1 for the use of La in this paper) as the target contour shape for its lack of the final tonal target.[6] The target APs in the first experiment differed in the numbers of syllables, four and five syllables, respectively. Since the accentual peak, i.e., +H, occurs either on the second or the third syllable, the presence of the tonal targets immediately after the +H might cause earlier peak alignment in the four syllable phrase than otherwise. By opting for a contour shape that lacks the final tonal target, it was intended to reduce the influence of the tonal space on the peak alignment.

The difference in the numbers of syllables and the syllable structure of the target phrases is due to the conflict among the controlling factors in the experiments; morpheme boundary and focus type. The material in the initial experiment was constructed primarily considering the factors implicated in Welby (2003, see also 1.2). The French factor is interpreted as the presence of the second phonological word or the morpheme boundary in Korean. This indicates that the initial phonological word of the two-word AP must not contain more than one morpheme, if we are to investigate the effect of the second phonological word on the AP peak alignment free from the influence of the morpheme boundary.

At the same time, the first word has to be longer than two syllables in order to demonstrate the effect of phonological word on the peak alignment. It should be noted that we assume that the second AP syllable is the default location of the accentual peak and that the presence of the second phonological word causes the peak to be aligned at the word boundary (see 2.1 and the Assumptions). If the first phonological word of the two-word AP contains only two syllables, according to the assumption, the peaks in both the one-word and the two-word APs would be placed equally in the second AP syllable and the effect of the second phonological word cannot be identified. Therefore, the first phonological word in the two-word AP has to be a single morpheme and contain three syllables or more, if the influence of the second phonological word is to be investigated. Therefore, for instance, /mi.ra.n -i/ in [[mi.ra.n -i]$_{PW}$ [ʌn.ni]$_{PW}$]$_{AP}$ 'Miran' (a girl's name)-suffix 'elder sister' (meaning 'elder sister Miran') cannot be employed as the first phonological word of the target two word AP, as it contains the suffix /-i/.

Korean morphemes as long as three syllables are not common. Furthermore, the 'one morpheme' condition restricts the use of Sino-Korean words, which constitute up to approximately 50~60% of the Korean vocabulary (Sohn 2001). In theory, individual syllables of Sino-Korean words can be regarded as separate morphemes as well as the whole or parts of the words. For that

[6]Also, the speakers employed L+HLa far more consistently than L+HL+Ha, for instance. Targeting L+HL+Ha frequently resulted in the use of other contour shapes, notably, L+HHa and L+HLa.

reason, by nature, morphemes and morpheme boundaries are not clear in Sino-Korean words and different speakers may interpret the morphological structure of the identical word differently. For instance, /tsa.doŋ.tɕ^{h}a / 'automobile, car' may be interpreted as one morpheme. It may also be interpreted as /tsa.doŋ. -tɕ^{h}a/, since /tsa.doŋ/ 'automatic' and /tɕ^{h}a/ 'car' are frequently used on their own, too (cf. /tsa.doŋ.mun/ 'automatic door' possibly as /tsa.doŋ. -mun/ and / kjʌŋ.tɕ^{h}al.tɕ^{h}a/ 'police (patrol) car' as /kjʌŋ.tɕ^{h}al -tɕ^{h}a /).

On the other hand, the one word AP may contain a morpheme of any length as well as clitics, as it is assumed that the peak is aligned with the second syllable regardless of the morphological structure of the phonological word/ AP. As a matter of fact, if the morpheme in the one word AP is 2½ syllable long and the peak is aligned later than the second syllable, but earlier than in the two word AP which contains a three syllable morpheme, the earlier peak alignment may be attributed to the shorter length of the morpheme in the one word AP.

Whereas the restrictions on the two word phrase require a long phrase for Korean, the test factor focus type allows only a limited use of particles, which essentially shortens the length of the one word phrase. Particles in Korean are all bound morphemes, and they specify and/or emphasise the grammatical functions (e.g., case) and/or the meaning (by adding a particular meaning) of the item which they are attached to. The target sentences have to be restricted in the use of particles in order to be employed as the answers to both the broad and narrow focus inducing questions. For instance, the target one word phrase in (2) can be made into a five syllable phrase by adding a particle /-ɾo/[7] 'to', i.e., /mi.ra.n-i.-ne. -ɾo/ instead of /mi.ra.n -i.-ne/, so that the number of syllables in both one and two word phrases are equal. However, in that case, the one word sentence can be used only as the answer to the question that specifically asks for the direction and, consequently, induces narrow focus on the target phrase.

In constructing the test material, the priority was given to the target phrase with two phonological words and the factors, such as, the morphological structure and the length of the initial phonological word, and the segmental content of the target phrase (for smooth F0 contours for the measurements), over the syllable structure and the number of syllables in the target APs (see 2.1.1). It was a decision which is based less on the experimental studies, but more on the empirical observation made over the years building a speech database for a speech synthesis system and, partially, on the intuition as a Seoul Korean native speaker. This decision may even be considered reckless. However, it is important to remember that the accentual peak is located, without exception, in the second or the third AP syllable, no matter how long an AP is. This suggests that, when an AP is four syllables or longer, AP length should not affect the accentual peak alignment in L+HLa which lacks the tonal targets

[7] /-ɾo/ simply adds directionality to the meaning of the lexical item it is attached to and may actually be translated as any preposition (or, in certain cases, adverb) in English that indicates direction, e.g. /wi. -ɾo/ 'up'-'to' meaning 'upward' or 'up'.

in the two final syllables (cf. L+HL+Ha). It should be also noted that the two target APs in the second experiment have four syllables each with L+HLa contour shape. Nonetheless, the peak is aligned with the second syllable in one AP and the third in the other (see 2.2.3). This supports the assumption that the different AP lengths should not affect the peak alignment in the first experiment. Also, the syllable structure (of the second AP syllable, at the very least) should not affect the accentual peak alignment, as is clearly indicated by the result of the second experiment. The target phrases differ in the syllable structure of the second AP syllable; one has CV and the other CVC, respectively. If the syllable structure affects the peak alignment, the peak should be aligned later in the phrase with the phonologically short CV syllable. That is, assuming that the accentual H peak is aligned relative to the second syllable, the peak should be aligned later in the CV syllable and earlier in the CVC. On the contrary, the result shows that the peak is not even located in the second syllable in the long CVC phrase, but is aligned early in the third syllable (see 2.2.3). This strongly suggests that the syllable structure does not affect the accentual peak alignment. Nonetheless, admittedly, the number of syllables and the syllable structure should have been better controlled.

2.1 Experiment 1

In this experiment, following assumptions are made and tested. We assume that the peak alignment is affected by the AP structure (i.e., the number of phonological words in an AP), sentence length (i.e., the presence or the absence of an AP before the target AP) and focus type. We assume that the accentual H tone is associated with the second AP syllable and the accentual peak is aligned with the syllable as default. However, when an AP contains two phonological words (or more), the peak is aligned at the end of the first word, because the stress in the second phonological word (see 1.1.2.) attracts the peak to be aligned in the stressed syllable, but the alignment is restrained by the phonological word boundary. We also assumed that the alignment of the accentual peak is affected by the presence/absence of the preceding AP and its presence causes later peak alignment in the following target AP. That is, the accentual peak is earlier in short sentences where the target AP is in utterance initial position than in long sentences where the AP is preceded by another AP. In addition, we look into focus intonation by hypothesising that narrow focus in Korean is marked by later peak alignment and higher tonal scaling than in broad focus.

Assumptions

1. The accentual peak is aligned with the default second AP syllable. However, when an AP contains more than one phonological word, the peak is aligned at the end of the first word (i.e., the beginning of the second word) due to the

influence of the word boundary and the following stressed syllable in the second word (see 1.1.2).
2. The accentual peak is aligned later, when there is a preceding AP.
3. Narrow focus is marked by later peak alignment and higher tonal scaling.

2.1.1 Description of the Experiment

Material The material consists of sets of casual conversational style questions and answers. The questions were constructed to induce different focus types, broad or narrow focus, in the answers. To induce broad focus, questions such as, 'what happened?' and 'what's new?' are more commonly used in the focus investigations. However, in this study 'What did you do?' was used instead, since the questions require a subject and the thematic subject particle, / -(i)ga/ in the answers, which puts emphasis on the subject. The questions in the material should induce focus on the target phrase, the first phrase of the sentences (2) and (3), regardless of the focus type. The phrase is marked by bold face in the gloss in (2∼6).

The answers in the material contain the target base sentences shown in (2) and (3); I've been to Miran's (home) and I saw 'Princess Aurora' (the title of a movie). The expected accentual phrasing is represented with square brackets and '$_{AP}$', e.g., [mi.ɾa.n -i.ne]$_{AP}$, in the transcription. The sentences differ mainly in the structure of the target first phrase; the initial phrase contains one phonological word in (2) and two words in (3), respectively. The one phonological word phrase consists of a content morpheme (opposed to the largely functional clitics) /mi.ɾan/ 'Miran (a girl's name)' and a suffix -i.ne '(someone's) home, family', meaning 'Miran's home'. The morpheme originally contains two syllables, however, it is produced with two and half syllables due to resyllabification. On the other hand, the two word phrase contains /o.ɾo.ɾa/'Aurora' and /koŋ.dʑu/[8] 'princess', that is, 'Princess Aurora' (the title of a well-known movie). Phonological words are marked with square brackets and '$_{PW}$', e.g., [o.ɾo.ɾa]$_{PW}$, in the transcription.

(2) [[mi.ɾa.n -i.ne]$_{PW}$] $_{AP}$ [ka.s'ʌ.s'ʌ.jo] $_{AP}$
'**Miran**'-'**home**' 'went'
I've been to Miran's.

(3) [[o.ɾo.ɾa]$_{PW}$ [goŋ.dʑ]$_{PW}$] $_{AP}$ [pwa.s'ʌ.jo] $_{AP}$
'**Aurora**' '**princess**' 'saw'
I saw 'Princess Aurora.

The target phrases differ in the number of syllables and the syllable structure (particularly of the initial syllable). This is due to the constraint imposed by the AP structure and focus type (see the beginning of 2. for more

[8]The lenis voicess velar plosive in the initial syllable (i.e. the onset consonant) of /goŋ.dʑu/ becomes voiced in /o.ɾo.ɾa.goŋ.dʑu/, as it is in intervocalic position.

detailed reasons). Recent research directs that the alignment of the accentual peak in Korean may be affected by the prosodic or morphological structure of an AP, that is, either the second phonological word in an AP or the morpheme boundary in a phonological word. To investigate if the presence of the second phonological word affect the peak alignment, the first phonological word in the two-word AP must not contain morpheme boundaries and has to be a morpheme with three syllables. Furthermore, employing the identical sentence in the broad and narrow focus investigation restricts the use of particles. The target phrases would have contained the identical number of syllables, if it had been possible to add a particle to the one-word phrase. However, particles specify the grammatical functions and the meanings of the lexical items that they attach to. Adding a particle to the one-word phrase would have prevented using the identical sentence in the investigation of the focus type. To minimise the possible influence of the different AP lengths, AP contours with final tonal targets, e.g., L+HL+Ha, were avoided, as the presence of the final tonal targets may affect the peak alignment. Instead, L+HLa was chosen as a target AP contour for the measurements (see below **Measurements**).

The location of the target phrases was varied between the sentence initial and medial positions by placing a phrase at the beginning of the base sentences. The target one word phrase (marked in bold face in the gloss) was preceded by sim.si.mɛ.sʌ '(I was) bored' in the long one-word sentence, (4).

(4)	[sim.si.mɛ.sʌ]$_{AP}$	[[mi.ɾa.n -i.ne]$_{PW}$]$_{AP}$	[ka.s'ʌ.s'ʌ.jo]$_{AP}$
	'*bored*'	'**Miran**'- '**home**'	'went'
	I was bored, so I've been to Miran's.		

(5)	[jʌ.dza.tɕʰin.gu.-ɾaŋ]$_{AP}$	[[o.ɾo.ɾa]$_{PW}$	[goŋ.dʑu]$_{PW}$]$_{AP}$	[pwa.s'ʌ.jo]$_{AP}$
	'*girlfriend*' - '*with*'	'**Aurora**'	'**princess**'	'saw'
	With my girlfriend I saw 'Princess Aurora'.			

(6)	[ʌ.dze.-nɯn]$_{AP}$	[[o.ɾo.ɾa]$_{PW}$	[goŋ.dʑu]$_{PW}$]$_{AP}$	[pwa.s'ʌ.jo]$_{AP}$
	'<u>yesterday</u>' - particle	'**Aurora**'	'**princess**'	'saw'
	Yesterday I saw 'Princess Aurora'.			

The target two word phrase was preceded by different phrases, /jʌ.dza.tɕʰin.gu.-ɾaŋ/ 'with (my) girlfriend'[9] in the broad focus sentence (5) and /ʌ.dze.-nɯn/ 'yesterday' in the narrow focus sentence (6). The dialogues containing the long two word sentences are shown in (7) and (8) and the target two word phrase is marked in bold face in the gloss (see the appendix for the full material). In the long broad focus sentence (A in (7)), 'with (my) girlfriend' is new information, as it is newly introduced into the conversation. In the long narrow

[9]This was replaced with 'with (my) boyfriend' for female speakers.

focus sentence (A in (8)), on the other hand, ‘yesterday’ is given information, as it was already mentioned in the question. Therefore, unlike the long one word sentences, the focus type distinction is aided by the information status of the sentence initial phrase in the long two word sentences. If the narrow focus intonation display distinct characteristics in the long two word sentences from the long one word sentences, the distinction may be attributed to the information status differences in the initial phrase.

(7) TWO WORD - LONG BROAD FOCUS

(Situation: On Monday, during a coffee break at work, you are having a chat with a colleague/friend.)

Q: What did you do at the weekend?
A: [jʌ.dza.tɕʰin.gu.-ɾaŋ]$_{AP}$ [[o.ɾo.ɾa]$_{PW}$ [goŋ.dʑu]$_{PW}$]$_{AP}$ [pwa.s’ʌ.jo]$_{AP}$
‘girlfriend’ - ‘with’ ‘Aurora’ ‘princess’ ‘saw’
With my girlfriend I saw ‘Princess Aurora’.

(8) TWO WORD - LONG NARROW FOCUS

(Situation: On Monday, during a coffee break at work, you are having a chat with a colleague/friend.)

Q: What movie did you see yesterday?
A: [ʌ.dze. -nɯn]$_{AP}$ [[o.ɾo.ɾa]$_{PW}$ [goŋ.dʑu]$_{PW}$]$_{AP}$ [pwa.s’ʌ.jo]$_{AP}$
‘yesterday’ - particle ‘**Aurora**’ ‘**princess**’ ‘saw’
Yesterday I saw ‘Princess Aurora’.

Speakers and recording Six native speakers of Seoul Korean (four male and two female) in their late 20s and 30s participated. They were all residents of Germany at the time of experiment and the duration of their stay ranged from six months to four years. They all reported to speak Korean daily.

The recording was made in the sound attenuated booth at IfL-Phonetics. The target sentences were embedded in eight dialogues (see the appendix). They were quasi-randomly ordered with fillers and presented on the cards with the participant’s lines highlighted. The author took up the role of the second speaker and asked the focus inducing questions. The participants were instructed to ‘answer’ the questions with the provided sentences rather than to read out. Other than that no further instructions were given. Approximately 10 repetitions for each condition[10] per speaker ($10 \times 8 \times 6$) were recorded directly on to a computer disk in 16 bit at the sampling rate of 44100 Hz.

Prosodic analysis We expected that short sentences (2∼3) would be produced typically with two APs and long sentences (4∼6) with three. We also expected L+HLa and LLa pitch contours for the APs in the target base sentences. The

[10]There were total of eight conditions, two conditions for each of the three control factors; Number of phonological words × AP location × focus type ($2 \times 2 \times 2$).

majority of the utterances were produced with the anticipated phrasing and AP contours. Some speakers, however, one of the female participants (LSE) in particular, occasionally produced the long ‘Miran’ sentence (4) with two APs, when it was in narrow focus. These were excluded from further analysis. This provided 471 utterances for labelling and analysis.

Measurements Using Praat (Boersma 1992), the beginning and the end of each AP and syllable were labelled. The F0 maximum in the target AP was located and the distance from the beginning of the second syllable, which is hypothesised to be associated with the AP initial H tone, to the F0 peak was measured (represented in Fig. 4). The measured distance was then divided by the duration of the second syllable and represented as a ratio of the second syllable for the comparison between different target sentences and speakers (see Table 1 for the normalised peak distance). The calculated value smaller than 1 indicates that the peak occurs in the second syllable and the value bigger than 1 indicates that the peak is located in the third syllable. For instance, the normalised peak distance 0.75 indicates that the peak is found about three quarters into the second syllable. On the other hand, 1.02 indicates that the peak is found at the very beginning of the third syllable.

For tonal scaling, F0 values were measured at the vowel centre of the assumed L tone syllables in the AP initial and the IP penultimate positions as well as the peak F0 value (Fig. 5). In addition, F0 values were extracted at the vowel centres of the syllables /goŋ/ (in (3), (5) and (6)) and /ne/ (in (2) and (4)) to investigate the influence of the AP structure on the focus intonation. /goŋ/ starts the second phonological word of the two word AP [[o.ɾo.ɾa]$_{\mathrm{PW}}$[goŋ.dʑu]$_{\mathrm{PW}}$]$_{\mathrm{AP}}$ ‘Princess Aurora’ and /ne/ is in the corresponding location in the one word AP [[mi.ɾa.n -i.ne] $_{\mathrm{PW}}$]$_{\mathrm{AP}}$ ‘Miran’s home’. Consequently, F0 was measured at the following locations:

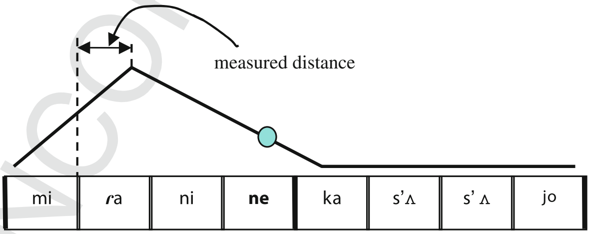


Fig. 4 Measuring of the AP peak distance. The line with a grey circle is the schematic representation of the AP pitch contours, L+HLa and LLa, of the utterance (2). The grey circle represents the location that corresponds to the vowel centre of /goŋ/ (see Fig. 5) in the utterance (the corresponding syllable is in bold face). Syllables are represented with squares and the AP boundaries with thick lines. The distance was measured from the beginning of the second syllable /ɾa/ (in bold italics) of the target AP to the F0 peak

This figure will be printed in b/w

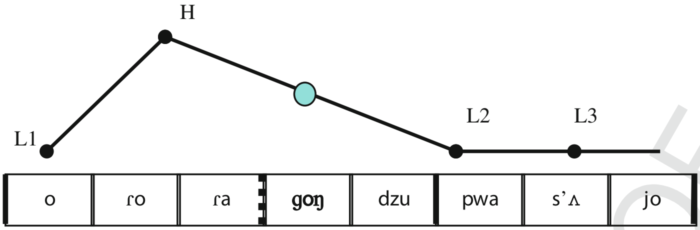


Fig. 5 Locations of F0 extraction. F0 values were extracted at the vowel midpoint of the assumed L tone syllables (in bold italics) and at the F0 maximum in the target AP. The locations of the F0 measurements are represented with L1, L2, L3 and H on the schematised pitch contour. In addition, F0 was measured at the vowel mid section of the syllable /goŋ/ (in bold face) which is the first syllable of the second phonological word of the target AP. This is represented with the grey circle on the contour. The phonological word boundary in the AP is indicated with a broken line

- at the F0 peak of the target AP (e.g., H in Fig. 5)
- at the vowel centre of the assumed L tone syllables L1, L2 and L3 (in bold face below in the phonetic transcription. See also in Fig. 5);

mi, ka, s'ʌ	in one word sentence	[**mi**.ɾa.n -i.ne]$_{AP}$	[**ka**.s'ʌ.**s'ʌ**.jo]$_{AP}$
o, pa, s'ʌ	in two word sentence	[**o**.ɾo.ɾa.goŋ.dʑu]$_{AP}$	[**pa**.**s'ʌ**.jo]$_{AP}$

- at the vowel centre of the initial syllable of the second phonological word in (3), (5) and (6) and the corresponding syllable in (2) and (4) (indicated with grey circles in Fig. 4 and Fig. 5);

  /ne/ and /goŋ/

The F0 values of L1, L2, L3 and H were normalised in semitone for the comparison of the intonation in broad and narrow focus sentences. The reference value was the mean F0 value of the L1 measurements in the long one-word and two-word sentences in broad focus, i.e., /mi/ and /o/ in long broad focus sentences. It was calculated for individual speakers (see Table 2). On the other hand, the syllables /ne/ and /goŋ/ were scaled against the initial syllables of the APs that they belong to and their reference values differed in each category and individual speakers (see Table 3). The syllable /ne/ was scaled against /mi/, which is the AP and phonological word initial syllable. The syllable /goŋ/, the initial syllable of the second phonological word in the AP [[o.ɾo.ɾa]$_{PW}$ [goŋ.dʑu]$_{PW}$]$_{AP}$], was scaled against the initial syllable of the first phonological word, /o/, which is, at the same time, the AP initial syllable.

2.1.2 Results

Peak alignment The measured peak distance (Fig. 4) was normalised in terms of the second syllable and represented as a ratio of the second syllable (see Table 1). The normalised values were subjected to a repeated-measures ANOVA with the factors NO. OF WORDS, SENTENCE LENGTH and FOCUS TYPE. There is a significant main effect of NO. OF WORDS ($F(1, 5)=8.021$, $p<0.05$) on the peak alignment. The peak is aligned earlier in one-word APs than in two-word APs. There is also a significant main effect of SENTENCE LENGTH ($F(1, 5)=7.879$, $p<0.05$) and the accentual peak is aligned earlier in short sentences than in long sentences. That is, the peak is earlier when the AP is in the utterance initial position and the presence of the preceding AP causes later peak alignment. However, no effect of FOCUS TYPE was reported.

There is an interaction SENTENCE LENGTH × FOCUS TYPE ($F(1, 5)=10.847$, $p<0.05$). This indicates that the FOCUS TYPE has a different effect on different sentence lengths. The comparison of the mean peak location shows that narrow focus causes earlier alignment in short sentences. The two-way interaction indicates that the early alignment in short narrow focus sentences is statistically significant. However, the interaction NO. OF WORDS × SENTENCE LENGTH × FOCUS TYPE ($F(1, 5)=7.298$, $p<0.05$) indicates that this is actually due to the early peak alignment in the short narrow focus two-word sentences. The three-way interaction graph in Fig. 6 shows that the accentual peak is aligned later in narrow focus than in broad focus with one exception; the peak is aligned earlier in the narrowly focused short two-word sentences. The three-way interaction indicates that the earlier peak alignment is significant at $p<0.05$ in the narrowly focused short two-word sentences.

Scaling of the L tones Since the distance between L2 and L3 was different in the the one-word and the two-word sentences, the one-word and the two-word sentences were subjected to a repeated-measures ANOVA separately with factors SENTENCE LENGTH, FOCUS TYPE and L-TONE LOCATION to understand the L tone scaling better. In one-word sentences, only the main effect of

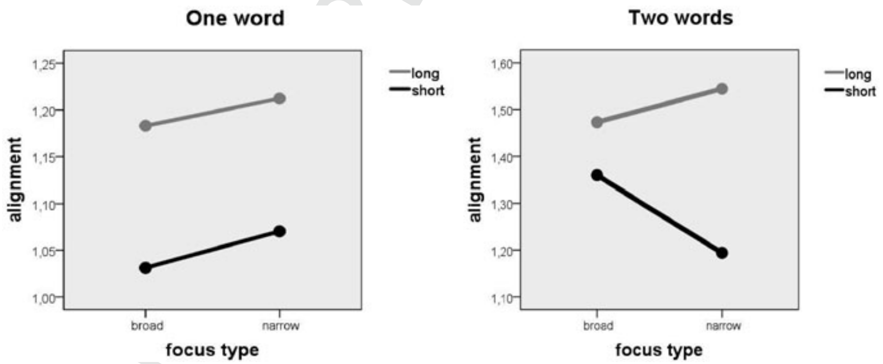


Fig. 6 Interaction graphs of NO. OF WORDS × SENTENCE LENGTH × FOCUS TYPE. The interaction NO. OF WORDS × SENTENCE LENGTH × FOCUS TYPE is presented in separate graphs according to the number of phonological words in an AP. In one-word APs, the peak alignment is later in narrow focus regardless of the sentence length, i.e., the presence/absence of the preceding AP. However, in short two-word sentences, the peak is aligned earlier in narrow focus than in broad focus

Table 1 Normalised peak distance. The table presents the peak distance for each speaker in each category. The measured peak distance (see Fig. 4) was normalised by dividing it with the duration of the second AP syllable. The mean and the standard error (in brackets) is presented for each individual in the bottom row

	oneword		(Miran)		twowords		(Aurora)	
	long		short		long		short	
Speaker	broad	narrow	broad	narrow	broad	narrow	broad	narrow
M1	1.110	1.161	0.899	0.983	1.387	1.337	1.171	0.852
M2	0.929	0.973	0.575	0.777	1.357	1.624	0.854	0.817
M3	1.381	1.510	1.459	1.332	1.426	1.460	1.341	1.133
M4	1.245	1.188	1.194	1.277	1.398	1.649	1.609	1.500
F1	1.086	0.956	0.858	0.861	1.654	1.629	1.553	1.413
F2	1.345	1.487	1.202	1.191	1.614	1.568	1.629	1.449
Mean	1.183 (0.07)	1.212 (0.1)	1.031 (0.13)	1.070 (0.09)	1.473 (0.05)	1.545 (0.05)	1.360 (0.12)	1.194 (0.13)

L-TONE LOCATION was reported significant ($F(2, 10) = 3.603$, $p<0.001$). L-TONE LOCATION interacted with SENTENCE LENGTH ($F(2, 10) = 4.153$, $p<0.05$). The scaling difference among L tones is bigger in short sentences than in long sentences and this is largely due to the high L1 in short sentences. The L-TONE LOCATION × SENTENCE LENGTH interaction indicates that higher L1 scaling in short sentences is significant. L-TONE LOCATION interacted with FOCUS TYPE ($F(2, 10) = 6.575$, $p<0.05$), too. The L tones in narrow focus sentences are scaled higher than in broad focus. The interaction plot (the left panel in Fig. 7) shows that the L tones in broad focus sentences fall in equal steps, that is, the difference between L1 and L2 is approximately the same as L2 and L3. However, in narrow focus they fall steeper between L2 and L3 and L3 is scaled at the similar level in both broad and narrow foci. The L-TONE LOCATION × FOCUS TYPE interaction indicates that the higher scaling of L1 and L2 is significant in narrow focus one-word sentences.

In the two-word sentences, too, only the main effect of L-TONE LOCATION ($F(2, 10)=39.431$, $p<0.001$) was significant. FOCUS TYPE was not significant, only very marginally ($p = 0.05$), and the L tones are scaled higher in narrow focus. The interaction effect of SENTENCE LENGTH × FOCUS TYPE was reported ($F(1, 5)= 13.587$, $p<0.05$). Narrow focus raises the L tone scaling in long sentences, whereas it barely affects the L tones in short sentences. The two-way interaction indicates that the higher tonal scaling in the long narrow focus sentences is significant. There were interaction effects of L-TONE LOCATION × SENTENCE LENGTH ($F(1.032, 5.161) =15.952$, $p<0.05$, Greenhouse-Geisser corrected) and L-TONE LOCATION × FOCUS TYPE ($F(2, 10) =12.513$, $p<0.01$). L1 is scaled higher in the short sentences than in the long sentences and in the narrow focus sentences than in the broad focus. On the other hand, L2 and L3 were affected little by the sentence length and the focus type and remained constant (see the right panel of Fig. 7 for the interaction plot L-TONE LOCATION × FOCUS TYPE). The interactions L-TONE LOCATION × SENTENCE LENGTH and L-TONE LOCATION × FOCUS TYPE indicate that the higher scaling of L1 is statistically significant in the short and the narrow focus sentences, respectively.

Scaling of the AP peak A three-way ANOVA (repeated-measures) with factors NO. OF WORDS, SENTENCE LENGTH and FOCUS TYPE shows that there are main effects of SENTENCE LENGTH ($F(1, 5)=9.522$, $p<0.05$) and FOCUS TYPE ($F(1, 5)=58.959$, $p<0.01$). The peak is scaled higher ($F(1, 48)=9.928$, $p<0.01$) in short sentences than in long sentences and in narrow focus than in broad focus. The interaction of NO. OF WORDS × FOCUS TYPE is reported ($F(1, 5)=7.708$, $p<0.05$). The effect of narrow focus is bigger in the one-word sentences and the peak is scaled higher in the one-word sentences than in the two-word sentences. There was also a three-way interaction of NO. OF WORDS × SENTENCE LENGTH × FOCUS TYPE ($F(1, 5)=26.105$, $p<0.01$). This indicates that sentence length affected NO. OF WORDS × FOCUS TYPE differently. The interaction graph (Fig. 8) shows that the effect of narrow focus was weaker in two-word sentences than in one-word sentences (that is, the interaction effect of NO. OF WORDS × FOCUS TYPE), because the peak scaling in short two-word sentences was not affected by narrow focus. Contrary to the short two-word sentences, the peak in the long two-word sentences is affected by narrow focus and scaled higher in narrow focus. The three-way interaction indicates that the higher peak scaling in long two-word narrow focus sentences is statistically significant.

Scaling of the syllables /ne/ and /ɡoŋ/ The syllable /ne/ was scaled against its phonological word initial syllable /mi/. The syllable /ɡoŋ/, which is the initial syllable of the second phonological word in the AP [[o.ɾo.ɾa]$_{\mathrm{PW}}$[ɡoŋ.dʑu]$_{\mathrm{PW}}$]$_{\mathrm{AP}}$, was scaled against the initial syllable of the first phonological word, /o/. It should be noted that the reference syllables /mi/ and /o/ are both the initial syllables of the APs that the target syllables /ne/ and /ɡoŋ/ belong to, and the results only reflect the relative pitch heights of the target syllables within the APs.

We ran two separate ANOVAs (repeated-measures, factors SENTENCE LENGTH and FOCUS TYPE) on /ne/ and /ɡoŋ/, respectively. No significant main or interaction effects is reported in the scaling of /ne/. Even though FOCUS TYPE does not have a significant main effect, the p-value is only marginally high with $p=0.53$ and the syllable is higher in narrow focus. It is highly likely that FOCUS TYPE is reported significant, had the data been bigger.

On the other hand, main effects of SENTENCE LENGTH on the syllable /ɡoŋ/ is reported ($F(1,5)=7.867$, $p<0.05$) and the syllable is scaled higher in long sentences. The interaction SENTENCE LENGTH × FOCUS TYPE is significant ($F(1,5)=52.490$, $p<0.01$), indicating that the effect of FOCUS TYPE is different in long and short sentences. The interaction graph in Fig. 8 shows that FOCUS TYPE has a contrasting effect according to the sentence length. In the long sentences, the syllable /ɡoŋ/ is scaled higher in narrow focus than in broad focus. However, in the short sentences, the syllable is scaled lower in narrow focus and the scaling difference between broad and narrow focus is bigger than in long sentences (Fig. 8). The two-way interaction indicates that the low scaling of /ɡoŋ/ in the short narrow focus sentences is significant.

Table 2 Scaling of the L tone syllables in each category. The mean is in semitone (see also the appendix for the reference Hz values and the L and H tone scaling of the individual speakers)

No. of words	length	focus type	L-tone location	Mean (st)	Std. error
one word (Miran)	long	broad	L1	0.43	0.22
			L2	−0.91	0.48
			L3	−2.85	0.42
		narrow	L1	1.04	0.28
			L2	−0.24	0.47
			L3	−2.63	0.40
	short	broad	L1	1.47	0.22
			L2	−0.94	0.81
			L3	−3.06	0.49
		narrow	L1	2.21	0.17
			L2	0.93	0.77
			L3	−2.66	0.45
two words (Aurora)	long	broad	L1	−0.49	0.24
			L2	−2.23	0.31
			L3	−2.94	0.47
		narrow	L1	0.41	0.30
			L2	−1.86	0.30
			L3	−2.72	0.44
	short	broad	L1	1.50	0.41
			L2	−1.71	0.67
			L3	−2.84	0.67
		narrow	L1	2.32	0.34
			L2	−2.14	0.61
			L3	−3.13	0.67

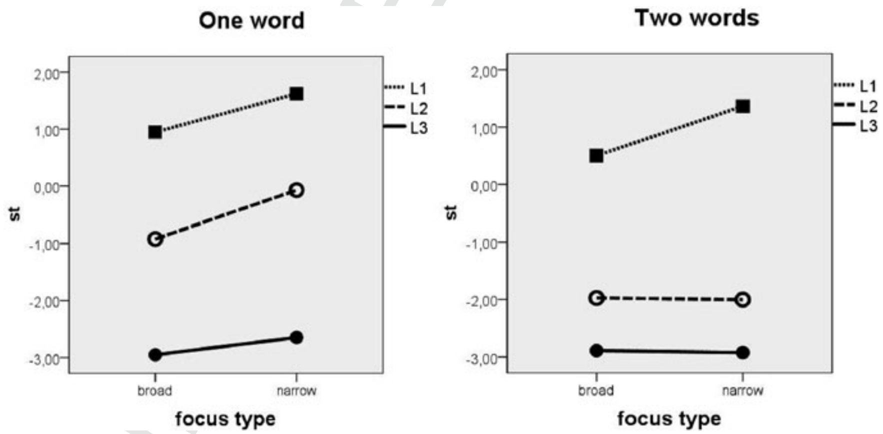


Fig. 7 Interaction graphs of L-TONE LOCATION × FOCUS TYPE in the one-word (on the left) and the two-word sentences (on the right). The graphs show that the interaction effect of L-TONE LOCATION × FOCUS TYPE is very distinct in the one-word and the two-word sentences. In the one-word sentences, narrow focus raises the scaling of L1 and L2. On the other hand, in the two-word sentences, only L1 is scaled higher in narrow focus

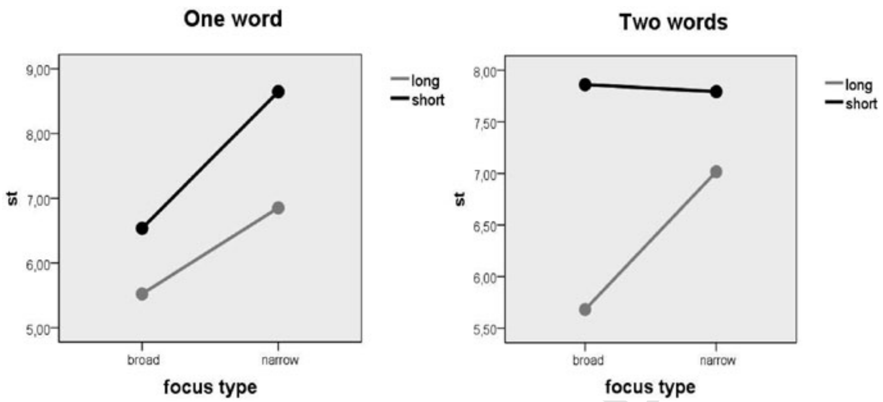


Fig. 8 Interaction graphs of no. of words × sentence length × focus type in the accentual peak scaling. Narrow focus raises the peak scaling and the peak is higher in narrow focus with the exception of short two-word sentences. In short two-word sentences, the peak is barely affected by narrow focus and scaled approximately the same in broad and narrow focus sentences

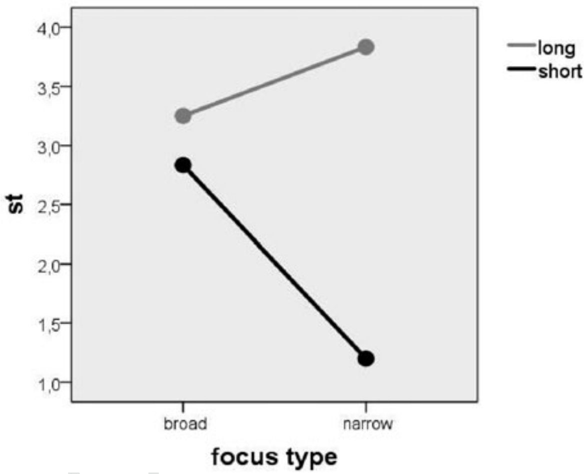


Fig. 9 Interaction of SENTENCE LENGTH and FOCUS TYPE in the scaling of the syllable /ɡoŋ/ in the two-word sentences. The gray bar represents the scaling in the long sentences and the dark bar in the short sentences. The graph shows that, whereas in long sentences /ɡoŋ/ is higher in narrow focus than in broad focus, in short sentences /ɡoŋ/ is scaled lower in narrow focus than in broad focus

2.1.3 Discussion

Accentual peak alignment The alignment results indicate that the alignment of the accentual peak is, as a whole, affected by the test factors; the number of words in the target APs, sentence length and focus type. The number of phonological words in the target APs affected the peak alignment and the peak is aligned significantly earlier in the one-word APs than in the two-word APs. However, contrary to the assumption, the peak is not placed in different syllables, but identically in the third AP syllable.

It was explained earlier that the alignment of the AP peak is similar to that of LHi (AP initial rise) in French. Analogously to French, we assumed that the AP peak alignment in Korean is affected by either the number of phonological words in an AP or the location of a morpheme boundary. It was also assumed that the accentual peak is placed in the final syllable of the first word near the word boundary in two-word APs. For that reason, the first phonological word in the target two-word AP has to be a three syllable morpheme in order to investigate the effect of the AP structure (i.e., the number of phonological words) on the peak alignment without the influence of a morpheme boundary. On the other hand, it was assumed that the peak is located in the second syllable in one-word APs and the target one-word AP was allowed to contain a morpheme boundary. As a result, the target one-word AP contains the initial morpheme that is 2½ syllables long and the target two-word AP contains the initial morpheme that is three syllables long. That is, the morpheme boundary is located in the third syllable in both the target APs, but it is earlier in the one-word AP than in the two-word AP. The alignment results show that the accentual peak occurs equally in the third syllable in the one-word and the two-word APs alike. Yet, the peak is aligned earlier in the one-word AP where the initial morpheme is shorter and this is statistically significant. The results suggest that the morpheme boundary location affects the accentual peak alignment rather than the number of phonological words.

The results indicate that sentence length affected the accentual peak alignment, too, and the peak is aligned later when the target AP is preceded by another AP. Narrow focus generally causes later peak alignment (see Fig. 6). Nonetheless, the statistical analysis shows that the effect of narrow focus is significant only in the short two-word sentences, where the peak is aligned earlier. (see **Narrow focus intonation** for further discussion.)

It seems that the information status of the preceding AP has little effect on the peak alignment of the target AP. Rather, it is the presence of the preceding AP itself that is significant. In the long one-word sentences, the identical sentence was used as the answer to the broad and the narrow focus inducing questions. However, in the long two-word sentences, the target APs were preceded by different phrases and the broad and narrow focus distinction was supported by the information status of the phrases; in the broad focus sentence, the preceding AP carried new information and in the narrow focus sentences, it carried given information. Nevertheless, the long two-word sentences show the identical alignment pattern as in the long one-word sentences without any distinct effect of focus type.

Accentual peak alignment in the third AP syllable It should be noted that the peak distance was normalised in terms of the second syllable duration, as it was assumed that the peak is aligned relative to the second AP syllable which the accentual H tone is assumed to be associated with. Since the accentual peak in both the target APs is located in the third syllable, the accentual peak location was newly represented as a ratio of the third AP syllable and it was investigated if the peak alignment still displays the identical characteristics and patterns. A (repeated-measures) ANOVA was applied to the new peak distance with the identical factors; NO. OF WORDS, SENTENCE LENGTH and FOCUS TYPE. A significant main effect of SENTENCE LENGTH was reported ($F(1,5)=25.695$, $p<0.01$) and the peak is aligned later in the long sentences than in the short sentences. There was an interaction effect of NO. OF WORDS $\times$ FOCUS TYPE ($F(1,5)=14.150$, $p<0.05$). The comparison of the means indicates that the peak is aligned later in the two-word sentences than in the one-word sentences (see Table 3) and the later peak alignment is significant only in broad focus. There was also an interaction of NO. OF WORDS $\times$ SENTENCE LENGTH $\times$ FOCUS TYPE ($F(1,5)=8.875$, $p<0.05$). The three-way interaction shows the identical patterns as Fig. 6. In the short two-word sentences alone, the peak is aligned earlier in narrow focus, while it is later in the other categories. The three-way interaction indicates that the earlier peak alignment caused by narrow focus is statistically significant in the short two-word sentences.

The alignment results are largely similar, even though different syllables were used as a reference in the normalisation. SENTENCE LENGTH and the interaction of NO. OF WORDS $\times$ SENTENCE LENGTH $\times$ FOCUS TYPE are consistently significant in the same way. The accentual peak is aligned significantly later in the long sentences. It is aligned significantly earlier in the short two-word sentence in narrow focus.

On the other hand, unlike the previous statistical analysis, no main effect of NO. OF WORDS is reported, when the peak distance is normalised with the third AP syllable duration. However, it should be noted that the two-way interaction NO. OF WORDS $\times$ FOCUS TYPE indicates that later peak alignment in the two-word APs is significant in the broad focus sentences. In addition, the three-way interaction NO. OF WORDS $\times$ SENTENCE LENGTH $\times$ FOCUS TYPE suggests that the peak latency in the two-word AP is not significant in the narrow focus sentences because of the early peak alignment in the short two-word sentences. Table 3 shows the peak location in each category represented as a ratio of the third AP syllable together with the peak latency in the two-word sentences (i.e., the alignment difference between the one-word and the two-word sentences). It shows that the peak is aligned later in the two-word sentences than in the one-word sentences, with the exception of the short two-word sentences in narrow focus. Contrary to the others, narrow focus causes earlier peak alignment in the short two-word sentences. The alignment is so early that it obscures the peak latency effect of NO. OF WORDS in the narrow focus sentences as a whole, even though the peak is aligned later in the long sentences. This suggests that the peak latency in the two-word sentences is significant and the effect of NO. OF WORDS should be significant. That is, as a matter of fact, the location of a morpheme boundary should be significant.

Table 3 The alignment of the accentual peak represented as a ratio of the third AP syllable. The table also shows the peak latency in the two-word sentences. Note that, with the exception of the short two-word sentences in narrow focus, the peak is aligned earlier in the one-word sentences. When the short two-word sentences are in narrow focus, however, the accentual peak is aligned earlier than in the corresponding one-word sentences

length	focus type	One word (Miran) Mean (Std. error)	Two words (Aurora) Mean (Std. error)	peak latency in two-words
long	broad	0.263 (0.101)	0.405 (0.059)	0.142
	narrow	0.291 (0.135)	0.477 (0.056)	0.186
short	broad	0.033 (0.182)	0.309 (0.113)	0.276
	narrow	0.095 (0.133)	0.020 (0.013)	–0.075

Narrow focus intonation Narrow focus also has different effects on the scaling of tones in the one-word and the two-word sentences, as well as the syllables /ne/ and /ɡoŋ/. The results in the L tone scaling show that the interaction L-TONE LOCATION × FOCUS TYPE affects the one-word and the two-word sentences distinctly, even though it is the interaction of the identical factors. Narrow focus causes higher scaling of L1 and L2 in the one-word sentences, whereas it raises only the scaling of L1 in the two-word sentences (compare the interaction graphs in Fig. 7).

Furthermore, the scaling of the AP peak displays different characteristics in the short two-word sentences. The peak scaling is not affected by narrow focus in the short two-word sentences and remains constant. On the contrary, the AP peak is strongly affected by narrow focus in the long two-word sentences, as well as in the one-word sentences, and scaled significantly higher in narrow focus. The effect of narrow focus is also manifested very differently on the scaling of the /ne/ and /ɡoŋ/. /ne/ in the one-word APs is not affected by focus type. The scaling of /ɡoŋ/, the initial syllable of the second phonological word in the two-word AP, is affected by narrow focus and the effect of narrow focus is strikingly different in the long and the short sentences; in the long sentences, the syllable is scaled higher in narrow focus, and in the short sentences, on the contrary, it is significantly lower in narrow focus. These effects of narrow focus are illustrated in Fig. 10 in comparison with broad focus.

The contours in Fig. 10 show that narrow focus intonation is characterised primarily by higher tonal scaling in the target AP rather than the difference in the accentual peak alignment. Narrow focus intonation is similar in one-word sentences and the long two-word sentence other than the scaling of L2. In one-word sentences, L2 is scaled higher in narrow focus, but in the long two-word sentence it remains constant. It seems that the L2 is scaled constant in the two-word sentences due to its proximity to the end of the utterance for being the antipenultimate syllable. Despite the scaling difference of L2, it should be noted that the narrow focusing strategy is essentially identical in the one-word and the long two-word sentences. The target APs are made prominent in terms of higher pitch and the following AP, on the contrary, is made less distinct with declining (in the one-word sentences) or relatively low pitch (in the two-word sentences).

On the other hand, in the short two-word sentence, only the first phonological word of the target AP is made prominent with high pitch. It should be reminded that the accentual peak is located in the final syllable of the first phonological word in the two-word APs. Higher L1 and the constant AP peak indicate that the first word is perceived higher in narrow focus than in broad focus. At the same time, the earlier alignment (and the constant scaling) of the accentual peak lowers the pitch on the following second word initial syllable /ɡoŋ/, creating a pitch level difference between the first and the second phonological words; the first word in the target AP is made prominent in terms of high pitch and the early peak alignment. This indicates that the alignment of the AP peak has an accentual function to lend prominence on to the AP. It should be noted that, even though only the first phonological word receives prominence in the short two-word sentence, it is the whole of the target AP that is focused. This suggests that the first word functions as a 'focus exponent' (Selkirk 1995) projecting focus on to the entire AP, when a two-word AP is narrowly focused in a short sentence.

In summary, the results show that narrow focus is manifested prosodically in two different ways. In long sentences (i.e., when the narrowly focused AP is preceded by another AP), the whole of the target AP is produced with higher pitch than in broad focus regardless of the number of words in the target AP. In short sentences (i.e., when the AP is placed utterance initially), the manifestation of narrow focus differs according to the AP structure. The whole of the target AP is produced with higher pitch, when the AP contains one phonological word, however, only the first phonological word is produced with higher initial pitch and earlier AP peak alignment, when the AP contains more than one word.

In the short sentences, the prosodic manifestations of narrow focus are particularly interesting in that they are related to two different AP structures. As a matter of fact, the different realisations of narrow focus reflect the strength relation of the component constituents that are immediately below AP, the

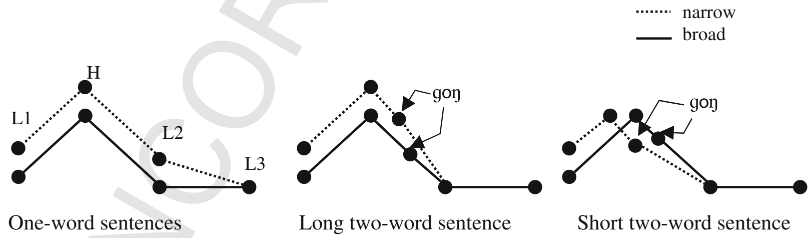


Fig. 10 Comparison of narrow and broad focus intonation. The effect of narrow focus (discussed above) is represented in the schematised pitch contours in comparison with broad focus. Separated circles represent statistically significant differences in the scaling (vertical) or in the alignment (horizontal). Note that, in the short two-word sentence, narrow focus intonation is marked with earlier peak alignment and the constant scaling of the accentual peak. Note also that the second phonological word initial syllable /ɡoŋ/ is scaled lower in narrow focus than in broad focus, whereas it is scaled higher in the long two word sentence

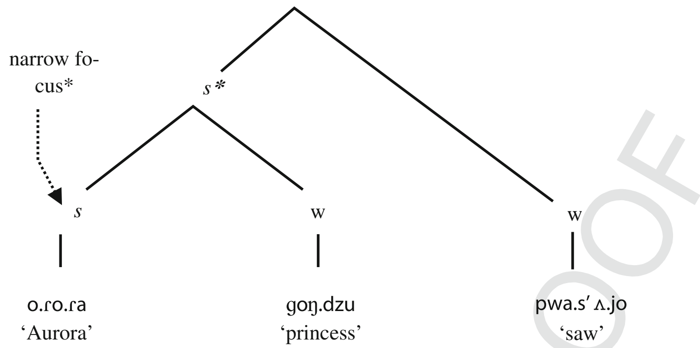


Fig. 11 Narrow focus and the strength relation in the short two-word sentence. The tree shows the hierarchical structure and the strength relations among the constituents in the sentence. It also illustrates that narrow focus affects the strong node of the AP 'Princess Aurora' (the nod is in italics) bringing prominence on to the entire nod (in italics with *)

phonological words. The two phonological words in the target two-word AP represent strong-weak relation and, when the AP is in narrow focus, only the strong node, i.e., first word, is made prominent in terms of prosody. In a one-word AP, the whole AP is made prominent, as it contains only one word which is a strong node. That is, narrow focus brings prominence onto the target AP by emphasising the strength relation of the word level constituents (see Fig. 11). Theoretically, an AP may contain unlimited number of phonological words, however, in practice, it rarely contains more than two words. An AP longer than that is difficult to produce and the words are usually produced as separate APs. Therefore, it may be said that one-word and two-word APs represent the prosodic structure of an AP and that the two different AP structures are reflected in the two different realisations of narrow focus. The fact that narrow focusing strategy differs according to the AP structure indicates that the difference in the AP structure is significant, and it is significant as it reflects the difference in the information structure. It should be remembered that the assumption in Lee's Korean intonation model (see 1.1.2.) is that a metrically strong word initiates an AP and the first phonological word is always metrically stronger than the following word(s), if there is any. By highlighting the first word in the two-word AP, the strong node becomes stronger relative to the weak nod and the weak nod becomes weaker relative to the stronger strong nod. By amplifying the strength relation in two-word APs, it is signalled that there is another piece of information on the same prosodic level, but the information is not as important as the preceding piece of information (or to start a new AP).

It should be pointed out that long narrow focus sentences are actually not very common in Korean. Korean is a 'situation language' with a very flexible word order. Given or shared information (or rather the information assumed to be shared) is not usually included in the utterances and the use the pro-forms, which contributes in yielding long sentences, is not very common. Word order is fairly variable. A verb or verbal phrase is restricted to be placed at the end of a sentence, however, it is also frequently omitted. In other words, the items that are deaccented or likely to be deaccented in languages like English are simply dropped in Korean. Long sentences are typically in broad focus and when in narrow focus, they are usually made shorter by omitting redundant information. For that reason, prosody is usually not required to make a focus type distinction in long sentences. On the other hand, in short sentences, focus type distinction is often made in terms of prosody. A narrow focus inducing question, such as, 'What movie did you see?', may be answered with (9)~(11).

(9)	[o.ɾo.ɾa.goŋ.dʑu]	'Princess Aurora' (the title of a movie).
(10)	[o.ɾo.ɾa.goŋ.dʑu -jo]	'Princess Aurora' - honorific particle
(11)	[o.ɾo.ɾa.goŋ.dʑu] [pwa.s'ʌ.jo]	'(I) saw 'Princess Aurora'

(9)~(11) reflect different degrees of politeness in casual and conversational speech in the order of (9) $<$ (10) $<$ (11). Higher degree of politeness and the use of honorifics than in (11) is very formal (almost too formal). Evidently, (9) and (10) can only be used in narrow focus context. However, (11) may also be used in the broad focus context (as in this experiment) and the use of prosody is required in order to make the focus type distinction (see the appendix for the full material). This indicates that prosodic distinction of focus type is more important in short sentences than in long sentences and the distinction has to be more distinctive. This explains the use of more complex focus strategy in short sentences, which reflects the AP structures.

At the beginning of the experiment, we made three assumptions regarding the alignment of the accentual peak in Korean APs. Firstly, we assumed that the alignment of the accentual peak is affected by the AP structure, i.e., the number of phonological words in an AP. It was assumed that the accentual peak is aligned with the second AP syllable as default, when an AP contains one phonological word. On the other hand, the peak is assumed to be aligned at the end of the first word (the beginning of the second word), when an AP contains more than one phonological word. Secondly, we assumed that the accentual peak is aligned earlier in short sentences where the target AP is in utterance initial position than in the long sentences where the AP is preceded by another AP. Thirdly, we hypothesised that narrow focus in Korean is marked by later peak alignment and higher tonal scaling than in broad focus.

The experiment results indicate that these assumptions are only partially right. They clearly show that sentence length affects the accentual peak alignment and the alignment is later in long sentences. They also show that narrow focus affects tonal scaling and the tones are scaled higher in narrow focus than in broad focus. However, focus type has very restricted effect on the peak alignment and narrow focus causes earlier peak alignment only in short two-word sentences. The number of phonological words, too, affects the peak alignment differently from the assumption. The peak is earlier in the one-word sentences than in the two-word sentences. However, the peak is not placed in the second syllable in the one-word AP, but in the third syllable same as the two-word AP. It seems that the peak alignment reflects the difference in the morpheme boundary location in the target APs. It should be remembered that both the target APs contain morpheme boundaries in the third syllable. However, the first morpheme is shorter by half syllable in the one-word AP than in the two-word AP; the first morpheme is 2½ syllables in the one-word AP and it is three syllables in the two-word AP. This suggests that the location of a morpheme boundary should affect the peak alignment and be responsible for the earlier peak alignment in the one-word AP rather than the number of phonological words. In the following Experiment 2, we investigate the hypothesis that the alignment of the accentual peak is affected by the location of a morpheme boundary and the presence/absence of the semantic content in the second morpheme.

2.2 Experiment 2

One of the hypotheses in Experiment 1 was that the accentual peak alignment is affected by the prosodic structure of an AP and that the alignment is earlier in the one-word AP than in the two-word AP. We assumed that the peak occurs in the default location of the second syllable in the one-word AP. We also assumed that the peak occurs in the third syllable in the two-word AP due to the influence of the word boundary and the following stressed syllable in the second word (see 1.1.2). The result of the experiment shows that the peak alignment is indeed earlier in the one-word AP than in the two-word AP. However, unlike the assumption, the peak in the one-word AP is not located in the second syllable, but in the third syllable as in the two-word AP.

The result indicates that the location of a morpheme boundary is likely to affect the peak alignment. It was explained earlier (see 1.2 and the beginning of 2.) that the hypotheses that AP structure affects the AP peak alignment is motivated by the findings in Welby (2003). She reports that the low F0 turning point created by the AP initial rise in French is aligned at the beginning of the content word when it is preceded by a function word. Otherwise, it occurs in the initial syllable. It was also explained that the function word-content word sequence is comparable to a Korean AP with two phonological words or with a content morpheme-functional morpheme (i.e., clitics) sequence and that this indicates that the AP peak alignment in Korean may be affected by the number of words or the location of a morpheme boundary. In Experiment 1, the first word of the two-word AP had to be a three syllable morpheme (without a morpheme boundary) in order to investigate the effect of the number of words on the peak alignment without the influence of a morpheme boundary. The one-word AP, on the other hand, was allowed to contain two and half syllable morpheme, as we assumed that the accentual peak occurs in the second syllable. That is, the morpheme boundary is located in the third syllable in the one-word and the two-word APs alike, however, it is earlier by half syllable in the one-word AP. Note that the accentual peak is aligned in the third syllable regardless of the number of words in the APs, but earlier in the one-word AP where the initial morpheme is shorter. This indicates that the accentual peak alignment is affected by the location of a morpheme boundary rather than the number of the phonological word. The French example also suggests that the peak alignment is affected by the presence/absence of semantic content in the second morpheme in Korean. It should be noted that the LHi alignment is varied by the presence/absence of the function word before a content word.

In Experiment 2, we tested the hypothesis that the alignment of the accentual peak is affected by the morpheme boundary location and the presence/absence of semantic content in the following morpheme. We constructed two noun phrases with a two syllable noun and the following two syllable particle. The particles differ in their functions; one was a functional case marker and the other an auxiliary particle with semantic importance. We assume that the morpheme boundary restricts the peak location, so that the peak is aligned at the end of the first morpheme. We also assume that this restriction may be overriden and the peak may occur in the first syllable of the second morpheme crossing the morpheme boundary, when the second morpheme is not functional (e.g., case markers) and has semantic content.

Assumptions

1. The accentual peak is aligned at the end of the first morpheme, when the following morpheme is functional.
2. The accentual peak is aligned in the first syllable of the second morpheme, when the second morpheme has semantic content.

2.2.1 Description of the Experiment

Material As in Experiment 1, the material consists of sets of casual conversational style questions and answers. The target sentences (12) and (13) were the answers to questions that were intended to induce narrow focus[11] on the target initial phrases. These phrases are marked with bold face in the gloss. The expected phrasing is represented with square brackets in the transcription.

The target phrases are made up of two morphemes, a two-syllable noun and a two-syllable particle (the morpheme boundary is represented with '-' in the

[11]It should be noted that, on strict terms, (12) and (13) differ in focus type (see the appendix for the full material). Unlike (12), the target initial phrase in (13) is in contrastive focus as it contrasts with /mɛ.il/ 'daily, everyday' in the question. Nonetheless, it is not likely that contrastive focus brings prominence on to /tsu.mal/ 'weekend' in /tsu.mal-.ma.da/ 'every weekend', as the contrast is in the semantics not the segments. The contrast is analogous to 'daily' and 'every weekend' in English.

transcription). The particles in (12) and (13) contrast in that contrary to the dative case marker /-e.ge/ in (12), the auxiliary particle /-ma.da/ in (13) adds the meaning ‘every’ to the preceding noun, e.g., /sa.ɾam -ma.da/ ‘person’-‘every’ meaning ‘every individual’. In Korean, particles are bound morphemes grouped into four subcategories (Huh 1993); case markers, auxiliary particles, conjunctional particles and special particles. With the exception of auxiliary particles, particles are mainly functional. Particularly, the use of the case markers is not obligatory and they may be dropped when the case information is retrievable from the context. For instance, the absence of the dative case marker /-e.ge/ in (12) does not affect the meaning of the sentence, since the sentence is an answer to the question ‘Who are you going to present this beautiful necklace?’ which specifically requires the information on the recipient and this makes it clear that the noun ‘Mina’ (a girl’s name) has to be dative. Auxiliary particles, on the other hand, do have semantic content and add various meanings to the preceding noun.[12] Without the auxiliary particle /-ma.da/ ‘every’, the initial phrase of (13) ‘every weekend’ becomes simply ‘weekend’.

(12) [mi.na. -e.ge] [tsul.k’ʌ.je.jo]
‘**Mina**’- $_{\text{dative case marker}}$ ‘will give/present’
(I) will present (it) to Mina.

(13) [tsu.mal. -ma.da]$_{AP}$ [man.na.jo]$_{AP}$
‘**weekend**’-‘**every**’ ‘meet’
(I) see (him/her) every weekend.

Speakers and recording 18 native speakers of Seoul Korean (six male and 12 female speakers) in their 20s and 30s participated. They were all residents of Germany at the time of the recording and the duration of their stay ranged from three months to five years.

The recording was made in the sound attenuated booth at IfL-Phonetics. The question and answer sets containing the target sentences were quasi-randomly ordered with filler question-answer sets and presented on the the cards with the participant’s lines highlighted. The author took up the role of the second speaker and asked the questions. The participants were instructed to ‘answer’ the questions with the provided sentences rather than to read out. Total of 216 utterances (6 repetitions × 2 target sentences × 18 speakers) were recorded directly on to a computer disk in 16 bit at the sampling rate of 44100 Hz.

Prosodic analysis The utterances were first checked for the irregularities in the F0 contours (for F0 measurements) and then subjected to prosodic analysis. They were produced with two APs with L+HLa and LLa contours,

[12]Huh (1993: 204) states ‘… [auxiliary particles] add special, refining meaning [to the noun which they attach to].’ (my translation)

respectively. However, three speakers (two male and one female) produced the target auxiliary particle phrase, notably, with L+HL+Ha. The utterances of the three speakers were not included in the investigation (see the beginning of 2. for the reasons). This left 188 utterances for further examination.

Peak distance measurement Syllables and phrases were marked using Praat and the distance from the beginning of the second syllable to the F0 peak in the target AP was measured (see Fig. 4)

2.2.2 Results

Case marker vs. auxiliary particle As in Experiment 1, the peak distance was normalised by dividing the measurements with the duration of the second AP syllable. The calculated mean is smaller than 1 in the case marker AP, indicating that the accentual peak occurs in the second syllable. On the other hand, the mean value is slightly bigger than 1 in the auxiliary particle AP, indicating that the peak occurs in the third syllable very near from the beginning crossing the morpheme boundary (see Table 4). A paired T-test indicates that the peak is aligned significantly earlier in the case marker phrase than in the auxiliary particle phrase ($t(15) = -7.387$, $p<0.001$).

2.2.3 Discussion

The result shows that the accentual peak is placed in different syllables in the two target APs. In the functional particle (i.e. the dative case marker) AP, the peak is aligned in the second syllable and in the auxiliary particle AP, it is aligned in the third syllable crossing the morpheme boundary. This suggests that the presence and the absence of semantic content in the second morpheme should influence the peak alignment.

It should be noted that the structure of the second syllable is different in the target APs. The syllable has CV structure in the case marker AP and CVC in the

Table 4 Mean peak location in the target APs. The table shows the peak location calculated in terms of the second and the third AP syllable duration. The values indicate that the accentual peak is located late in the second syllable in the dative case marker AP, whereas it is early in the third syllable in the auxiliary particle AP. Additionally, the peak alignment result from Experiment 1 is provided in the corresponding sentence length and focus type (short and narrow focus) for comparison (separated with an empty row). However, the alignment in the two-word AP is provided with that of the 'broad focus' sentence (marked with *) due to the exceptionally early peak alignment in the two-word AP in narrow focus (see Experiment 1)

AP types	σ2 normalised mean peak location (Std. error)	σ3 normalised mean peak location (Stdev.)
case marker	0.746(0.04)	-
auxiliary particle	1.068(0.05)	0.091(0.25)
one-word AP (Miran)	1.070(0.09)	0.095(0.33)
two-word AP (Aurora)*	1.360(0.12)	0.309(0.28)

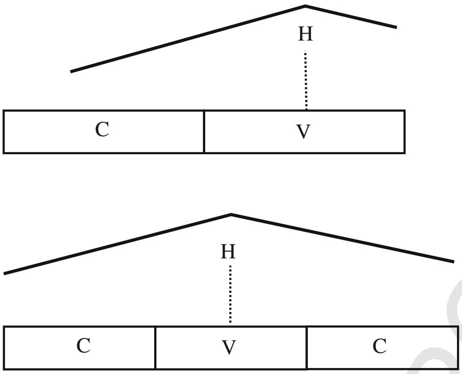


Fig. 12 Alignment of the H peak anchored to the vowel centre of a syllable. The figures illustrate the H peak alignment in a syllable with different structures. The peak is anchored to the centre of the vowel in both the illustrations. However, the peak is aligned relatively earlier in the CVC syllable than in the CV syllable

auxiliary particle AP. Theoretically, the structure of the second syllable is important, as the second AP syllable is assumed to be associated with the accentual H tone and the difference in the phonological length of the syllable may result in the difference in the accentual peak alignment. The second syllable in the 'weekend' phrase (average 180 ms) is actually longer than the 'Mina' phrase (average 174 ms) phonetically as well as phonologically. However, the result indicates that the syllable structure does not affect the peak alignment. If the phonological (or phonetic) length of the second syllable had affected the peak alignment, the alignment result would have been the opposite. That is, contrary to the result, the peak should have been aligned earlier in the 'weekend' AP than in the 'Mina' AP. The structure (or the length) of the second AP syllable affects the accentual peak alignment, if the accentual H peak is anchored to (or aligned with reference to) a specific point in the syllable. If, for instance, the AP peak is anchored to the centre of the vowel in the second AP syllable, the peak should be aligned relatively earlier in the CVC syllable (i.e., the 'weekend' AP) than in the CV syllable (the 'Mina' AP), as illustrated in Fig. 12. The fact that the peak is aligned later in the 'weekend' AP than the 'Mina' AP suggests that the structure of the second AP syllable does not affect the accentual peak alignment and that the AP peak is not aligned relative to the second syllable.

The experiment result is compared with that of Experiment 1 in the corresponding sentence length and focus type, i.e., the short narrow focus sentences (see Table 4). Only, the peak alignment in the two-word AP is provided with the peak distance measured in the short narrow focus sentences (marked with * in Table 4). This is due to the contrasting effects of narrow focus in Experiment 1. Narrow focus caused later peak alignment, with the exception of the short two-word AP. For that reason, it was reckoned that the peak distance should be provided with the measurement in the broad focus sentences in the two-word AP.

Table 4 compares the peak alignment in the APs that contain the initial morphemes of different lengths; the case marker and the auxiliary particle APs contain two syllable morphemes each, and the one-word 'Miran' and the two-word 'Aurora' APs from Experiment 1 contain 2½ and three syllable morphemes, respectively. It should be noted that the peak is aligned later in the order of case marker < auxiliary particle/one-word AP < two-word AP, reflecting the growing length of the AP initial morpheme. The different number of speakers in Experiment 1 and Experiment 2 does not allow any statistical tests. However, it should be noted that the peak alignment in the APs in Table 4 was already reported statistically significant in individual experiments, suggesting that the length of the AP initial morpheme should be significant. It is also important to note that the peak is aligned almost the same location at about 1.07 in the one-word AP (Experiment 1) and the auxiliary particle AP (Experiment 2). The statistical analysis indicates that the alignment is significantly earlier in the short one-word AP (with 2½ syllable morpheme) than in the two-word AP (with three syllable morpheme) in Experiment 1. It also indicates that the peak alignment is significantly earlier in the case marker AP (with 2 syllable initial morpheme) than in the auxiliary particle AP, which has the identical peak alignment as the short one-word AP. This suggests strongly that the peak alignment difference among the three categories of APs, the case marker, the one-word and the two-word APs should be significant and also that the accentual peak alignment is affected by the length of the AP initial morpheme, i.e., the location of a morpheme boundary.

The comparison indicates that the alignment of an accentual peak in Korean is affected by the location of a morpheme boundary. The occurrence of the peak is confined to the AP initial morpheme and the peak is aligned later as the length of the morpheme increases. The importance of the morpheme boundary location in the peak alignment suggests that speaker's interpretation or conception of a morpheme should influence the AP peak alignment and result in the alignment variation among different speakers. When the location of a morpheme boundary is not clear or may be conceived differently among different speakers, as it may with some compounds, foreign loan words or Chino-Korean words (e.g., /tsa.doŋ.tɕʰa/ 'automobile' or /tsa.doŋ.mun/ 'automatic door', see also the beginning of 2.), the accentual peak alignment may vary according to the location of the conceived morpheme boundary.

The comparison also suggests that, as we assumed, the peak alignment should be restrained by a morpheme boundary. At the same time, the alignment in the case marker and the auxiliary particle APs suggests that this constraint may be overridden by the high semantic weight of the following morpheme. It should be also noted that, as mentioned in 2.2.1, unlike the case marker phrase, the AP tone in the auxiliary particle phrase shows some variation and, most notably, L +H L+ Ha was observed. The final rise of the contour occurs in the particle /-ma.da/ ‘every’ making the particle perceptually prominent. This indicates that speakers may employ a different AP tone, as well as accentual peak placement, to bring prominence onto the semantically important part(s) of an AP.

3 General Discussion

In the Korean intonation model which provides the theoretical basis for this study, it is assumed that an AP is defined by the tonal pattern THLH (T= L or H). The phonetic variants of the AP tone pattern usually contains a peak either on the second or the third syllable, which is assumed to be the phonetic exponent of the initial H tone. In this study it was investigated if the following factors affect the peak alignment; the number of phonological words, sentence length (i.e., the presence/absence of the preceding AP), focus type, the location of a morpheme boundary and the presence of semantic content in the following morpheme.

The results indicate that the morpheme boundary affects the peak alignment systematically. The peak is aligned at the end of the first morpheme and its alignment becomes later as the length of the initial morpheme increases. On the other hand, the peak occurs in the second morpheme when the morpheme has semantic content. This indicates that the occurrence of the peak is restricted to the initial morpheme, however, this restriction may be overridden by the presence of the semantic content in the following morpheme. The influence of morpheme boundary on the accentual peak alignment suggests that the AP initial H tone should be associated with an edge of a morpheme. We may assume that the H is, as a rule, associated with the right edge of the first morpheme of an AP, but it gets associated with the left edge of the second morpheme when the second morpheme has semantic content.

It should be noted that the alignment of the AP peak varies within a very limited range of the second and the third AP syllables and, most importantly, the peak never occurs later than the third syllable. This indicates that there has to be yet another constraint which limits the occurrence of the peak to the beginning of an AP and that the accentual H tone is associated with an edge of a morpheme under this constraint. The constraint cannot be the accentual H tone’s association to the second AP syllable as assumed in Jun (1996, 2000). The experiment results show that the peak alignment is not affected by the phonetic or phonological duration of the second AP syllable, indicating that the peak is not aligned relative to the syllable. Theoretically, this implies that the accentual H tone is not associated with the second AP syllable.

Since the results do not support the assumption that the accentual H tone is associated with the second AP syllable, we propose that the accentual H

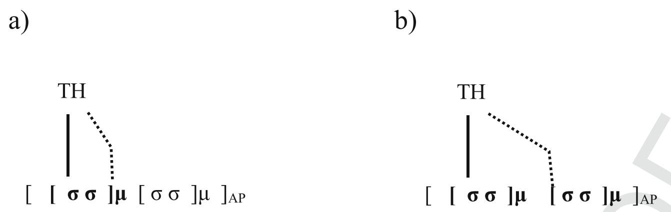


Fig. 13 The association of the two AP initial tones. (a) represent the association of the two AP initial tones TH, when the AP contains two morphemes and only the first morpheme has semantic content (content morpheme in bold face). The tones are associated with the AP initial syllable as a unit (represented with a solid line) and, separately from T, the accentual H gets associated simultaneously with the right edge of the first morpheme (represented with a dashed line). (b) represents the tonal association when the AP contains two content morphemes. The TH are associated with the AP initial syllable as in (a), however, the accentual H gets associated with the left edge of the second morpheme

tone should be analysed as a component tone of TH associated with the initial syllable of an AP. That is, similarly to a bitonal accent in English, the initial TH is analysed as a single tonal event. We also assume that the H tone is simultaneously associated with the right edge of the first morpheme in an AP. However, when the second morpheme has semantic content, it gets associated with the left edge of the morpheme. That is, the H tone is doubly associated with the AP initial syllable (as a component tone of TH) and an edge of a morpheme (see Fig. 13). However, this proposal is merely a tentative analysis and requires to be substantiated with further studies.

4 Summary and Conclusions

At the beginning, we hypothesised that the peak alignment in Korean AP is systematic and linguistically conditioned. The experiment results show that the alignment of the accentual peak is indeed systematic and the alignment is affected by the presence of a preceding AP, the location of a morpheme boundary and the presence of semantic content in the following morpheme. The alignment of the accentual peak is restricted by a morpheme boundary. The occurrence of the peak is confined to the AP initial morpheme and as the morpheme becomes longer, the peak is aligned later. This constraint is overridden by the semantic importance of the following morpheme, however. When the following morpheme has semantic content, the peak occurs in the following second morpheme. The result that the peak placement is limited to the content morphemes and affected by the presence of semantic content in a morpheme suggests that the peak placement should have accentual function to bring prominence.

The influence of the morpheme boundary on the accentual peak alignment suggests that the H tone is associated with an edge of a morpheme. At the same time, the fact that the variation in the peak alignment is restricted to the second and the third AP syllable suggests that there should be another constraint in the peak alignment which restricts the occurrence of the peak to the beginning of an AP. Since the results do not support the assumption that the accentual H tone is associated with the second AP syllable, we proposed to analyse the AP initial TH as a tonal unit similar to a bitonal accent in languages such as English. The TH is assumed to be associated with the AP initial syllable and, at the same time, the H is associated with an edge of a morpheme separate from the initial T (see Fig. 13).

Narrow focus does not affect the peak alignment, but in the short two-word AP. The accentual peak is aligned earlier in the short two-word AP. Narrow focus, as a whole, has distinct effect on the short two-word AP. In other APs, narrow focus makes the entire AP prominent by raising the scaling of tones in the target AP. On the other hand, in the short two-word AP, only the first phonological word is made prominent in terms of pitch. The AP initial L tone is scaled higher in all the target APs in narrow focus. However, in the short two-word AP, earlier peak alignment has an effect of lowing the scaling of the second word initial syllable, making the first word prominent. That is, the whole of the AP is made prominent by making the first word prominent in terms of pitch. This suggests that the peak alignment has accentual function to bring prominence onto a part (or whole) of an AP. It also suggests that the first phonological word in the short two-word AP functions as a focus exponent and projects focus on to the entire AP.

Acknowledgments I would like to thank Martine Grice for her insightful and invaluable discussions and advice. I would also like to thank the two anonymous reviewers for their thorough comments. My special thanks to Reinhold Greisbach for his help in statistics.

APPENDIX A

MATERIAL - EXPERIMENT 1

SHORT BROAD FOCUS – ONE WORD

Situation: You had an accident and was in a hospital. Since a few days you are at home recovering. A friend of yours came by to see you on a Saturday afternoon.

Q: What did you do all day today?
A: [mi.ɾa.n -i.ne] [ka.s’ʌ.s’ʌ.jo]
‘**Miran**’- ‘**home**’ ‘went’
→ I’ve been to Miran’s.

SHORT NARROW FOCUS – ONE WORD

Situation: You had an accident and was in a hospital. Since a few days you are at home recovering. You came just back home from a neighbour's, when a friend of yours came by to see you on a Saturday afternoon.

Q: We were worried! Where the hell have you been?
A: [mi.ɾa.n -i.ne] [ka.s'ʌ.s'ʌ.jo]

→ I've been to Miran's.

SHORT BROAD FOCUS – TWO WORD

Situation: On Monday, during a coffee break at work, you are having a chat with a colleague/friend.

Q: What did you do at the weekend?
A: [o.ɾo.ɾa. goŋ.dʑu] [pwa.s'ʌ.jo]
'**Aurora**' '**princess**' 'saw'

→ I saw 'Princess Aurora'.

SHORT NARROW FOCUS – TWO WORD

Situation: On Monday, during a coffee break at work, you are having a chat with a colleague/friend.

Q: What movie did you see?
A: [o.ɾo.ɾa. goŋ.dʑu] [pwa.s'ʌ.jo]

→ I saw 'Princess Aurora'.

LONG BROAD FOCUS – ONE WORD

Situation: You had an accident and was in a hospital. Since a few days you are at home recovering. A friend of yours came by to see you on a Saturday afternoon.

Q: What did you do all day today?
A: [sim.si.mɛ.sʌ] [mi.ɾa.n̤.i.ne] [ka.s'ʌ.s'ʌ.jo]
'*bored*' '**Miran**'- '**home**' 'went'

→ I was bored, so I've been to Miran's.

LONG NARROW FOCUS – ONE WORD

Situation: You had an accident and was in a hospital. Since a few days you are at home recovering. You came just back home from a neighbour's, when a friend of yours came by to see you on a Saturday afternoon.

Q: We were all worried! Where the hell have you been?
A: [sim.si.mɛ.sʌ] [mi.ɾa.n̩.i.ne] [ka.s'ʌ.s'ʌ.jo]
'bored' '**Miran**'- '**home**' 'went'

→ I was bored, so I've been to Miran's.

LONG BROAD FOCUS – TWO WORD

Situation: On Monday, during a coffee break at work, you are having a chat with a colleague/friend.

Q: What did you do at the weekend?
A: [jʌ.dza.tɕ^{h}in.gu̩. ɾaŋ] [o.ɾo.ɾa. ɡoŋ.dʑu] [pwa.s'ʌ.jo]
'girlfriend' - *'with'* '**Aurora**' '**princess**' 'saw'

→ With my girlfriend I saw 'Princess Aurora'.

LONG NARROW FOCUS – TWO WORD

Situation: On Monday, during a coffee break at work, you are having a chat with a colleague/friend.

Q: What movie did you see yesterday?
A: [ʌ.dze. nɯn] [o.ɾo.ɾa. ɡoŋ.dʑu] [pwa.s'ʌ.jo]
'yesterday' - particle '**Aurora**' '**princess**' 'saw'

→ Yesterday I saw 'Princess Aurora'.

MATERIAL - EXPERIMENT 2

Q: It's such a beautiful necklace! Who are you going to present it to?
A: [mi.na. -e.ɡe] [tsul.k'ʌ.je.jo]
'**Mina**'- dative case marker 'will give/present'

→ (I) will present (it) to Mina.

Q: You used to meet up with your boyfriend/girlfriend for lunch. Do you still see him/her *everyday?
A: [tsu.mal. -ma.da] [man.na.jo]
'**weekend**'- '**every**' 'meet'

→ (I) see (him/her) every weekend.

* [mɛ.il] is used for 'everyday'.

APPENDIX B

Reference F0 values for tone and syllable scaling in Experiment 1

Reference F0 values for tonal scaling

speaker	Ref. value (Hz)
M1	115.4
M2	90.4
M3	95.8
M4	146.0
F1	181.4
F2	204.6

Reference F0 values and scaling of /ne/ in one-word AP

	long broad focus			long narrow focus			short broad focus			short narrow focus		
Speaker	ref. value (Hz)	scaling in st	stdev	ref. value (Hz)	scaling in st	stdev	ref. value (Hz)	scaling in st	stdev	ref. value (Hz)	scaling in st	stdev
M1	114.8	1.08	0.06	116.3	1.12	0.04	127.4	1.10	0.04	135.5	1.22	0.09
M2	91.3	1.07	0.05	94.5	1.12	0.06	97.7	0.92	0.05	100.4	1.04	0.06
M3	97.7	1.08	0.06	99.3	1.16	0.09	106.5	1.22	0.13	109.2	1.24	0.11
M4	146.1	1.12	0.13	162.3	1.38	0.08	150.0	1.25	0.06	165.8	1.54	0.13
F1	194.5	1.09	0.09	201.3	1.14	0.05	202.8	1.04	0.06	211.4	1.15	0.16
F2	217.3	1.20	0.06	217.5	1.28	0.04	224.3	1.30	0.02	225.2	1.25	0.05

Reference F0 values and scaling of /ɡoŋ/ in two-word AP

	long broad focus			long narrow focus			short broad focus			short narrow focus		
speaker	ref. value (Hz)	scaling in st	stdev	ref. value (Hz)	scaling in st	stdev	ref. value (Hz)	scaling in st	stdev	ref. value (Hz)	scaling in st	stdev
M1	115.9	1.44	0.22	125.0	2.31	0.39	130.5	1.08	0.73	138.4	0.39	1.16
M2	89.5	2.04	0.50	94.5	2.23	0.59	97.6	−0.65	0.21	99.7	−1.43	1.14
M3	93.8	1.68	0.77	96.6	2.42	0.83	109.8	0.65	1.34	112.3	−0.61	0.89
M4	145.9	2.59	0.87	149.0	4.23	0.89	156.3	2.83	0.95	167.3	1.37	1.27
F1	168.2	6.76	0.62	173.5	5.42	1.59	207.0	5.56	0.94	216.3	2.51	1.27
F2	191.9	5.04	0.66	201.6	6.29	0.99	213.8	7.59	1.02	217.2	5.25	0.99

References

Baumann, S., J. Becker, M. Grice, and D. Muecke. 2007. Tonal and articulatory marking of focus in German, *Proceedings of the 16th ICPhS*, Saarbruecken, 1029–1032.

Beckman, M., and J. Pierrehumbert. 1986. Intonational Structure in Japanese and English, *Phonology Yearbook* 3: 255–309

Boersma, P. 1992. Praat: doing phonetics by computer, University of Amsterdam, http://www.fon.hum.uva.nl/praat/

Braun, B. 2006. Phonetics and phonology of thematic contrast in German. *Language and Speech* 49(4): 451–493.

Braun, B. 2007. Effects of dialect and context in the realisation of German prenuclear accents, *Proceedings of the 16th ICPhS*, Saarbruecken, 961–964.

Cho, T., and P. Keating. 2001. Articulatory Strengthening at the Onset of Prosodic Domains in Korean, *Journal of Phonetics* 28: 155–190.

Crystal, D. 1969. *Prosodic Systems and Intonation in English*. Cambridge: Cambridge University Press.

Face, T. 2001. Focus and early peak alignment in Spanish intonation. *Probus* 13: 223–246.

Frota, S. 2002. Tonal association and target alignment in European Portuguese nuclear falls. In Carlos Gussenhoven and Natasha Warner (eds.) *Laboratory Phonology 7*, 387–418. Berlin/New York: Mouton de Gruyter.

Huh, Ung. 1985. *Kugeo Eumunhak*, Seoul, Saemmunhwasa.

Huh, Ung. 1993. *Kugeohak, Seoul*, Saemmunhwasa.

Jang, T-Y. 2000. *Phonetics of Segmental F0 and Machine Recognition of Korean Speech*. Phd Thesis. University of Edinburgh.

Jun, Sun-Ah. 1996. *The Phonetics and Phonology of Korean Prosody*, New York: Garland.

Jun, Sun-Ah. 2000. K-ToBI (Korean ToBI) Labeling Conventions, http://www.humnet.ucla.edu/humnet/linguistics/people/jun/ktobi/K-tobi.html

Jun, Sun-Ah, and C. Fougeron. 2000. A Phonological Model of French Intonation. In Antonis Botinis (ed.) *Intonation: Analysis, Modeling and Technology*, 209–242, Kluwer Academic Publishers.

Laniran, Y. O. 1992. *Intonation in Tone Languages: The Phonetic Implementation of Tones in Yoruba*, Ph. D dissertation, Cornell University.

Lee, H.-B, and C. Seong. 1996. Experimental phonetic study of the syllable duration of Korean with respect to the positional effect, *Proceedings of the 4th International Conference on Spoken Language Processeing*, 1193–1196.

Lee, H.-Y. 1990. *The Structure of Korean Prosody*, Ph. D dissertation, University of London.

O'Connor, J. D., and G. Arnold. 1973. *Intonation of Colloquial English*, Longman

Pierrehumbert, J. 1980. *The Phonology and Phonetics of English Intonation*, Ph.D. dissertation, Massachusetts Institute of Technology.

Pierrehumbert, J., and M. Beckman. 1988. *Japanese Tone Structure*, Cambridge, MA: MIT Press.

Selkirk, E. O. 1984. *Phonology and Syntax: The Relation Between Sound and Structure*, Cambridge, MA: MIT Press.

Selkirk, E. O. 1995. Sentence Prosody: Intonation, Stress, and Phrasing. In John A. Goldsmith (ed.) *The Handbook of Phonological Theory*, 550-569. Cambridge, MA/ Oxford, UK: Blackwell.

Sohn, Ho-Min. 2001. *Korean Language*, Cambridge, MA: MIT Press.

Uhmann, S. 1990. *Fokusphonologie*, Niemeyer.

Welby, P. 2003. *The Slaying of Lady Mondegreen*, Ph.D dissertation, Ohio State University.

Yonsei Institute of Language and Information Studies 1998. *Yonsei Korean Dictionary* http://kordic.britannica.co.kr

The Perception of Negative Bias in Bari Italian Questions

Michelina Savino and Martine Grice

1 Conversational Moves and Intonation in Bari Italian

The present study has been motivated by results from a previous investigation on the relationship between pragmatic categories and their intonational marking in the Bari variety of Italian. These studies are based on the analysis of task-oriented dialogues, elicited by using a specially adapted version of the HCRC Map Task (Anderson et al. 1991). The Map Task is a non-linguistic task involving verbal cooperation between two participants (an Instruction Giver IG, and an Instruction Follower IF), each having a map. The task consists of reproducing as accurately as possible the route which is drawn on one of the maps onto the other map. The task is complicated by the fact that the two maps are not identical, as there are a number of differences in the presence and position of the landmarks across the two maps.

A specific dialogue structure coding scheme (Carletta et al. 1997) has been developed for the Map Task distinguishing three hierarchical levels of dialogue analysis, which are the following (from the highest to the lowest):

Transactions, 'which are subdialogues that accomplish one major step in the participants' plan for achieving the task' (Carletta et al. 1997: 14);

Conversational games, which 'embody the observation that, by and large, questions are followed by answers, statements by acceptance or denial, and so on' (Carletta et al. 1997: ibidem). Conversational games are also differentiated between *initiations* 'which set up a discourse expectation about what will follow' and *responses* 'which fulfil those expectations' (Carletta et al. 1997: ibidem);

M. Savino (✉)
Department of Psychology, University of Bari, Bari, Italy
e-mail: m.savino@psico.uniba.it

S. Frota et al. (eds.), *Prosodic Categories: Production, Perception and Comprehension*, Studies in Natural Language and Linguistic Theory 82,
DOI 10.1007/978-94-007-0137-3_8,

Conversational moves, which are ‘simply different kinds of initiations and responses according to their purpose’ (Carletta et al. 1997: ibidem). Conversational moves are therefore the building blocks of the whole dialogue structure: initiating moves are typically questions, instructions or explanations (QUERY, CHECK, INSTRUCT and EXPLAIN moves in the Map Task coding scheme), whereas response moves are replies to questions, clarifications, or backchanneling (REPLY, CLARIFY and ACKNOWLEDGE moves in the same coding scheme).

Our previous studies have investigated the relationship between a number of initiating and response conversational moves and their intonational marking (Grice and Savino 1995a; Grice et al. 1995; Savino 1997, 2000, 2001, 2004, Grice and Savino 1997, 2003a, 2003b, 2004). Most of them concentrate on the distinction between two different types of questioning moves, namely QUERY-YN, i.e. yes-no information-seeking questions, where the information sought is new (i.e. unknown by the speaker at the time of asking the question), and CHECK, i.e. confirmation-seeking questions, where information is already shared (given) (Carletta et al. 1997). Furthermore, in QUERIES there is not necessarily a bias as to the expected answer to the question, whereas CHECKS are biased towards a positive answer. Examples of one QUERY-YN (Example 1) and one prototypical CHECK (Example 2) found in the Bari Italian Map Task dialogues are in the following excerpts taken from Grice and Savino (2003b, 2004), where the target utterance is in boldface:

Example 1 (QUERY-YN):
IG: piega di nuovo verso destra
(*turn again towards the right*)
IF: sì
(*yes*)
IG: **a questo punto <eeh> hai un leone?**
(***at this point <uh> do you have a lion?)***
IF: <eeh> sì, sul margi+ cioè diciamo quasi a metà sulla destra
(<uh>yes at the edge that is let’s say almost halfway up on the right)

Example 2 (prototypical CHECK):
IG: continua continuando
(*continue by continuing*)
IF: verso il basso?
(*towards the bottom?)*
IG: no continuando verso sinistra
(*no continuing towards the left*)
IF: **<ah!> in tratto orizzontale**
(***<ah!> horizontally?***)
IG: sì sì in obliquo leggermente in obliquo sì
(*yes yes diagonally slightly diagonally yes*)

In our analysis, we also found utterances which, although they repeat an item, and thus appear to refer to given information, actually challenge what has been said. We could not analyse those utterances as CHECKS as the speaker is not asking for confirmation of shared information, but rather showing disbelief and challenging the interlocutor's assumption that information is shared (Tench, 1996). In the literature, these utterances have been referred to as 'echo questions' (*inter alia*, Cruttenden 1986) as they (partially) repeat what has been previously said by the interlocutor, and can also signal '[. . .] varying shades of incomprehension, doubt or surprise' (Bartels 1999: 158). This led to the proposal of a new category in the Map Task annotation scheme for describing this type of utterance in our Italian dialogues that we named OBJECT (as in objection) (Grice and Savino 1995a, 1997, 2003a, 2003b; Carletta et al. 1997). An example of OBJECT moves (in boldface) encountered in one of the Bari Italian Map Task dialogues is in the following excerpt (from Grice and Savino 2003b):

Example 3 (prototypical OBJECT):
IG: ce l'hai il ristorante Anima Mia?
(*do you have it, restaurant My Soul?*)
IF: **Anima MIA?!?**
(***My Soul?!?***)
IG: eh
(*yeah*)
IF: **ANIMA?!?**
(***SOUL?!?***)
IG: eh
(*yeah*)

An OBJECT type of question has a strong bias toward a negative answer, rather like a question with 'really' in English (Romero 2006; Romero and Hann 2004). Thus, if we take 'Anima mia' to be an elliptical question, its translation could be reformulated as follows: 'Did you really say Anima mia?' or 'Do you really have a restaurant with the name Anima mia?'.

In the Bari Italian Map Task dialogue sessions, participants were not told in advance that the two maps were different (neither were they told that they were identical, even though this is what participants assumed). Thus, OBJECT moves typically occur when a speaker makes this discovery, usually when a landmark is missing in one of the two maps or is placed differently in the two maps. In this case, an OBJECT move signals that the existence or position of a landmark cannot be verified. Whereas a CHECK move asks for confirmation that the proposition should be accepted and integrated into the common ground, an OBJECT move negates the proposition and indicates that it should *not* be integrated into (or should be even removed from) the assumed common ground. The meaning of this contour is also akin to Ward and Hirschberg's incredulity contour, L* + H L-H%, in which 'it *is* the case that the speaker believes a scale or scalar is *in*appropriate' (Ward and Hirschberg 1988: 515, italics in original).[1]

The OBJECT move has been further categorised in the analysis of Australian English Map Task dialogues, where Stirling et al. (2001) treated it as a type of ACKNOWLEDGE move. They describe OBJECT as 'a minimal negative response to a move indicating that it was understood but not accepted' (Stirling et al. 2001: 117) and thus do not treat it as a question at all. However, since OBJECTS do usually require a response, we continue to treat OBJECT as a type of question, albeit with a strong negative bias regarding the polarity of the propositional content (Romero 2006).

2 Intonational Marking of QUERY-YN, CHECK and OBJECT Conversational Moves

As mentioned above, the background work of the present paper is our previous investigation on the relationship between pragmatic categories and their intonational marking in the Bari variety of Italian, especially in the case of QUERY-YN, CHECK and OBJECT conversational moves. Some observations relevant for the present paper are described in the sections below.

2.1 QUERY-YN – CHECK Distinction

We have observed that, at the pragmatic level, the distinction between QUERY-YN and CHECK moves can be considered as continuous: between the two extremes, asking for new information and asking for confirmation about assumed given information, there can be different degrees of speaker's confidence as to whether information is new or not. Basing our pragmatic analysis on orthographic transcriptions of the dialogues, we found a large number of cases where information was textually given – i.e. already mentioned in the dialogue – but was assumed to be new by the interlocutor, typically because it was mentioned many turns before and therefore was not currently active in his/her consciousness (Chafe 1974). In those cases, we observed that some of the utterances classified as CHECKS on the basis of textual analysis might be in fact QUERIES from the point of view of the speaker for whom the information is inactive (Grice and Savino 2004: 14). In other cases, when information was mutually given or accessible (Chafe 1974), we observed in confirmation questions (CHECKS) degrees of speaker's confidence as to the givenness of information.

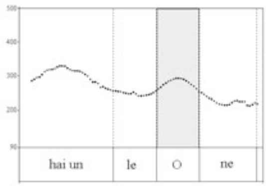
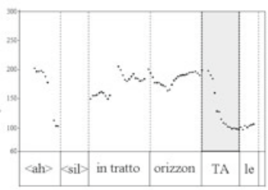
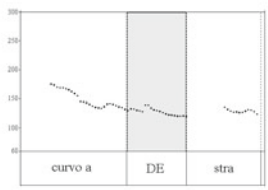
On the other hand, we observed that the intonational marking of such pragmatic gradience is discrete, as we found three different types of pitch

[1]However, it is not possible to compare directly, as Bari Italian lacks a counterpart with the same tonal analysis which has the meaning of uncertainty.

accent, each related to the degree of speakers' confidence that information is given, namely:

– a rising L + H* pitch accent for QUERY and for a sub-category of CHECKS we introduced as 'tentative CHECK', when the speaker confidence as to the correctness of inferred material (i.e. that information is shared) is very low (example in Fig. 1a in Table 1)

Table 1 Schematisation of the QUERY – CHECK distinction in Bari Italian: speakers use different pitch accent types according to the degree of confidence that information is shared. Fig. 1a shows the F0 contour of the rising L + H* pitch accent typical for Queries and tentative Checks (it is the Query 'hai un leOne?' '*do you have a lion?*' in Example 1), Fig. 1b that of the high falling H* + L pitch accent used for confident Checks ('ah, in orizzonTAle' '*ah, horizontally*' in Example 2), and Fig. 1c the F0 shape of the low falling H + L* pitch accent typical of very confident Checks ('e curvo a DEstra' '*and I have to curve to the right*')

Pragmatic Function	Intonational Marking *discrete (accent type)*
- QUERY - tentative CHECK **neutral to slightly positive bias**	L+H* L+H*  Fig. 1a
confident CHECK **positive bias**	H*+L  Fig. 1b
very confident CHECK **strong positive bias**	H+L*  Fig. 1c

– a high falling $H^* + L$ pitch accent for what we named 'confident CHECK' (example in Fig. 1b, Table 1)
– a low falling $H + L^*$ pitch accent for the third sub-category we called 'very confident CHECK' (example in Fig. 1c, Table 1). Note that when no context is provided, these utterances are indistinguishable from statements.

It is worth noting that intonationally this distinction is marked by what is analysed as three quite different pitch accents. The main difference in pitch contour between QUERIES and tentative CHECKS on the one hand, and confident CHECKS on the other is the presence of a dip before the H* peak, as can be observed in Fig. 1a. This dip, analysed as an L leading tone, has been shown to be crucial for the perception of a QUERY (Grice and Savino 1995b). Peak timing also appears to be contrastive for the confident vs. very confident CHECK distinction: in our data the peak is realised within the nuclear syllable ($H^* + L$) in the confident CHECKS, whereas it is before the nuclear syllable ($H + L^*$) in the very confident ones. However, further experimental work is needed to support these observations.

The schematisation discussed above is shown in Table 1.

2.2 QUERY-YN – OBJECT Distinction

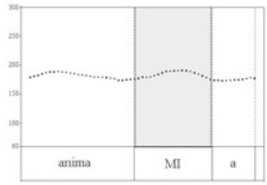
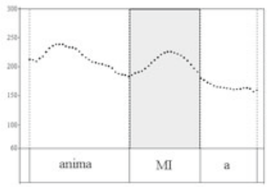
The QUERY – OBJECT distinction is marked in what might look like a gradient way, as both are characterised by the same rising $L + H^*$ pitch accent, although the peak is higher in OBJECT than in QUERY, as it is illustrated in Fig. 2a and Fig. 2b, Table 2.

Note also that such a difference cannot be captured in the current phonological descriptive framework, where the height of F0 peak is either high (H*) or downstepped (!H*) in relation to a previous peak (Ladd 1996), there being no provision for extra high in the standard analyses of Italian.

The idea of an 'extra high' peak was originally proposed by Pike (1945): in his description of English intonation he included four levels – from 1 to 4 – where level 4 corresponds to the 'extra high' (Overhigh) level. The concept of an Overhigh tone has been further entertained by Ladd (1994), suggesting that the F0 peak could be at the same height or higher (i.e. upstepped) with respect to a previous peak.

The distinction between QUERY, OBJECT and CHECK questions is made on the accented syllable, the edge tones being low for all three (L-L%). In QUERIES and CHECKS, the final boundary tone can also be high (H%), (Grice et al. 2005, also for further references), although its use appears to be related to aspects of speaking style, where high boundary tones are typically produced in reading, whereas L% tones prevail in spontaneous speech (Grice et al. 1997; Refice et al. 1997).

Table 2 Schematisation of the QUERY vs OBJECT distinction in Bari Italian: they share the same tonal structure but have a different peak height (more compressed in QUERIES than in OBJECTS). Fig. 2a shows a spontaneous rendition of the QUERY 'Anima MIa?' (*My Soul?*) by a female Bari Italian speaker, and Fig. 2b a spontaneus rendition of the OBJECT 'Anima MIa?!?' (*My Soul?!?*) by the same Bari Italian speaker. Note that at the current stage of our analysis, [high peak] is used as a notational variant of the L + H* pitch accent

Pragmatic Function	Intonational Marking *possibly gradient (peak height)*
QUERY asking for new information **neutral bias**	L+H*  Fig. 2a
OBJECT challenging assumed given information **negative bias**	L+H*[high peak]  Fig. 2b

Results from these production studies thus indicate that the intonational distinction between QUERY and OBJECT in Bari Italian relies mainly on the height of the peak on the nuclear accent.

Perceptual evidence that the role of peak height is not confined to paralinguistics (as claimed in traditional studies, see for example Bolinger 1989) has been gathered by a number of previous studies, demonstrating that listeners can also make linguistic use of such variations (Hirschberg and Ward 1992; Ladd and Morton 1997, replicated by Chen 2003 with reaction time measurements, Vanrell Bosch 2006a, 2006b, Borràs-Comes et al. 2010).

In marking the QUERY vs OBJECT distinction in Bari Italian, we cannot of course exclude the influence of other prosodic features – as suggested for example by Hirschberg and Ward (1992) for the incredulity vs uncertainty interpretation of the rise-fall-rise in American English. Further, it is not clear whether it is a matter of the height of an individual peak only, or of the whole utterance (as for example observed by Gili Fivela (2008b: 142) for OBJECT intonational contours in Pisa Italian, where 'a globally high prenuclear F0 stretch' is analysed as involving a high left boundary %H). However, in the present study we concentrate on peak height, which is equivalent to pitch range for the purposes of our experiment, since the utterance we used is short and contains only one H tone.

3 Research Question and Methodology

The aim of the present investigation is to obtain experimental evidence that indeed peak height variation plays a role in perceiving QUERY and OBJECT as two different types of question in Bari Italian. In order to verify whether listeners are able to reliably label an utterance as QUERY or OBJECT by listening to stimuli manipulated for peak height only, we carried out an identification task followed by a discrimination task. The first was a semantically motivated identification task with a binary choice (Savino and Grice 2007). In both tasks, we not only recorded listeners' responses but also measured Reaction Time. Reaction Time (henceforth RT) reflects a subject's uncertainty in making a decision (Pisoni and Tash 1974; Repp 1984), reflecting therefore the cognitive load involved in the decision making process: listeners are faster in labelling non-ambiguous stimuli and slower with ambiguous ones. Since our aim is also to explore listener confidence in the categorical interpretation of the stimuli as QUERY or OBJECT, we consider RT as a good measure of such confidence.

Recording RT in perceptual experiments involving the manipulation of F0 has already been used in previous studies (see Chen 2003, who pioneered this method for assessing intonational contrasts in English, Vanrell Bosch 2006b in Majorcan Catalan, Falé and Hub Faria 2006 in European Portuguese, Gili Fivela 2008a, 2008b in Pisa Italian and, more recently, Borràs-Comes et al. 2010 in Catalan). Results in identification tasks have shown that judging stimuli at the category boundary produced longer RT than stimuli at the extremes of the phonetic continuum. We expected to find a similar trend in our identification task.

In order to explore listeners' ability to discriminate between pairs of stimuli in the continuum, i.e. when they are not specifically asked to label them as QUERY or OBJECT, we carried out a discrimination task (performed directly after the identification task), also with RT measurements. We considered RT as a good measure of listeners confidence in this task too (but see results reported in Gili Fivela 2008b, pointing to a different evaluation of RT measurements in discrimination tasks). It is worth noting that we did not necessarily expect to find a discrimination peak in subject responses, as predicted by the Categorical Perception paradigm (Liberman et al. 1957; Repp 1984): the great majority of the previous studies using discrimination tasks in intonation failed to show a clear discrimination peak in responses (Ladd and Morton 1997; Remijsen and van Heuven 1999; Schneider and Lintfert 2003), calling into question the adequacy of such a paradigm in investigating contrasts in intonation. By measuring RT in the discrimination task we nonetheless expected to have shortest RT at/around the stimulus pairs corresponding to the category boundary in the identification task.

4 Identification Task

A semantically motivated identification task was designed, in which subjects were asked to judge a number of stimuli created along a phonetic continuum as belonging to one or the other of the two given pragmatic classes. This continuum was created by increasing and decreasing the F0 peak of a basic stimulus derived by a one accented naturally produced utterance with a medial pitch range. According to the classical Categorical Perception paradigm, if the curve of the function representing the percentage of responses for a category is S-shaped, first evidence of categorical interpretation will be provided.

4.1 Preparation of Stimuli

A trained female speaker of Bari Italian produced the utterance ‘a miLAno?’ (in Milan?) with an intended mid-way pitch accent between a QUERY and an OBJECT. [2] A stylised version of this utterance was used as the base stimulus for creating a phonetic continuum of 12 different versions, by systematically varying the peak height of the pitch accented syllable. Note that since we decided, as a first step, to consider only one F0 peak in determining the phonetic continuum, peak height and pitch range can be regarded as equivalent for the purpose of our experiment. Using one base utterance of this kind helped us in controlling the parameter we were concentrated on, that is peak height, avoiding the influence of other possible prosodic cues.

Starting from the base stimulus (F0 peak = 235.6 Hz), 4 stimuli were obtained by decreasing the peak height in 15 Hz steps, and 8 stimuli were produced by shifting upwards the peak height, also by 15 Hz steps. This procedure resulted in a continuum of 13 stimuli, as shown in Fig. 3, where the

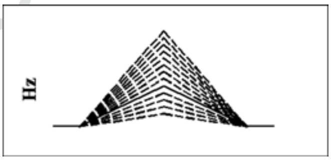


Fig. 3 Phonetic continuum of stimuli created for the perceptual test. The base stimulus (solid line) has an F0 peak of 235.6 Hz. Remaining stimuli (dashed lines) were produced by systematically decreasing and increasing the peak height by 15 Hz steps

[2] Thank you to David House for this suggestion.

base stimulus is represented by a continuous line. Note that the two extremes of our phonetic continuum, i.e the lowest and the highest peaks, were determined as follows: the lowest peak as the penultimate 15 Hz step before reaching the baseline, and the highest peak as the last acceptable one beyond which pitch manipulation produced distorted voice quality.

F0 manipulation was performed by using the PSOLA resynthesis programme implemented in the PRAAT software package for speech analysis and resynthesis (Boersma and Weenink 1999).

4.2 Presentation of Stimuli

Stimuli were presented in 4 blocks of 13, each block preceded by 2 warning tones and followed by 10 seconds of silence. Each stimulus was preceded by 1 warning tone and followed by 4 seconds silence for answering. The first block was treated as a training set and was not taken into account in the statistical analysis. Before starting the task, informants were given written instructions presenting two different possible contexts for the utterance 'a Milano' to be produced. The two dialogues as they appear in the instruction sheets, along with the English translation, are given below.

DIALOGUE 1: A: 'La prossima riunione dei G8 si farà in Italia' (*The next G8 meeting will take place in Italy*) B: **'<u>A Milano?</u>'** (*in Milan?)* A: 'Sì, a Milano' (*yes, in Milan*)
DIALOGUE 2: A: 'Stamattina a Milano c'erano 45 gradi' (*This morning there was 45 degrees in Milan*) B: '**<u>A Milano?!?</u>** Ma cosa dici, non e' possibile!' (*in Milan?!? What are you saying, it isn't possible*) A: 'Sì, a Milano, ti dico' (*yes, in Milan, I'm telling you*)

Explanations of the two contexts were also provided, as follows:

In the first dialogue participant B is asking a question aiming at obtaining some information, typically by a negative or positive answer. In this case, B wants to know whether the next G8 meeting will take place in Milan or not;
In the second dialogue, on the other hand, B is not simply asking for a piece of information, but with that question is doubting the preceding statement expressed by A. In this specific case, B does not believe it is possible that the temperature in Milan could reach 45 degrees.

Stimuli were presented on a computer over headphones, and informants were asked to judge whether each stimulus produced by speaker B occurred in dialogue 1 or dialogue 2, by pressing the appropriate button on the computer keyboard ('1' for dialogue 1 and '2' for dialogue 2).

To facilitate the recalling of the button function, the following text was shown on the computer screen during the whole session:

1 = **'A Milano?'**

2 = **'A Milano?!?'**

Subjects were asked to answer as quickly as possible, but in any case not before they had listened to the utterance. They were also warned they had a maximum of 4 seconds of time available for answering, and that it was not possible to skip the answer for any of the utterances presented in the sequence. The experiment was carried out in a quiet laboratory and the experimenter was always present but did not interfere during the task.

The perceptual experiment was implemented by using the E-Prime software tool which allows Reaction Time recording. In this case, RT was the time from stimulus onset until the subject presses a button on the computer keyboard to answer.

4.3 Informants

13 Bari Italian listeners (aged between 20 and 45) participated in the experiment on a voluntary basis. They were all recruited among staff and students of the two local universities (mostly coming from the Faculty of Engineering), and none of them had a background in linguistics, phonetics or prosody.

4.4 Results

4.4.1 Responses

The percentage of response agreement (Fig. 4) shows the typical S-shaped curve of categorical interpretation, confirmed by a probit analysis ($R^2 = 0.998$). Following Chen (2003), the location of category boundary was calculated by performing a linear regression analysis on 'Query' response frequencies corresponding to stimuli 5, 6, 7, 8, 9 in the phonetic continuum. In this case, the linear regression analysis shows that the mentioned response frequencies can be a reliable predictor of the category boundary ($R^2 = 0.989$; $F = 291.08$; $p = 0.0004$).

The linear regression equation (Y = a1 * X + a0) in our case is the following:

$$Y = -21.2821 * X + 195.1282$$

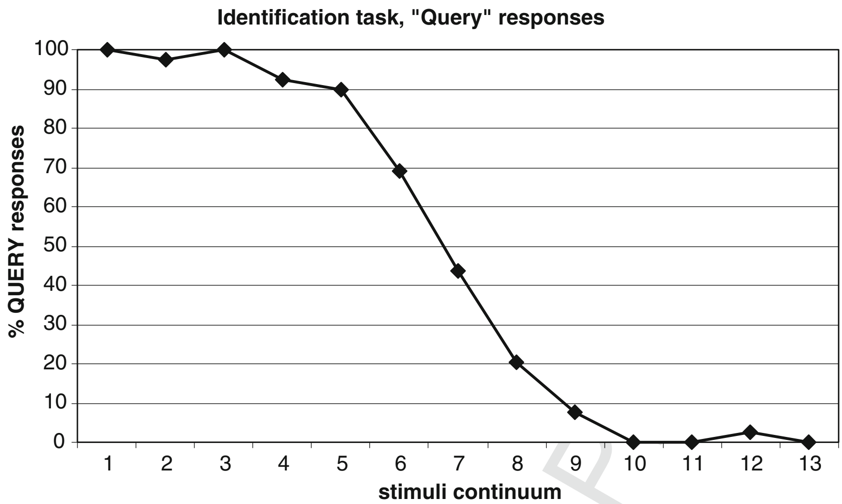


Fig. 4 Percentage of judgements as Query in the semantically motivated identification task. The location of the category boundary (calculated by linear regression analysis, and confirmed by probit analysis) corresponds to stimulus 7 in the phonetic continuum

For Y = 50, we obtain X = 6.82, where X is the location of the category boundary. In practice, the boundary predicted by this formula corresponds to stimulus 7 (6.82). The same result was obtained by calculating the category boundary by the mentioned probit logistic regression (6.76).

We also looked at the consistency of judgements across listeners: Fig. 5 shows the category boundary determined by each listener on the phonetic continuum, which was calculated by applying the same method we used for pooled data (linear regression analysis) described above. Results show that all listeners indicated the category boundary consistently around stimulus 7.

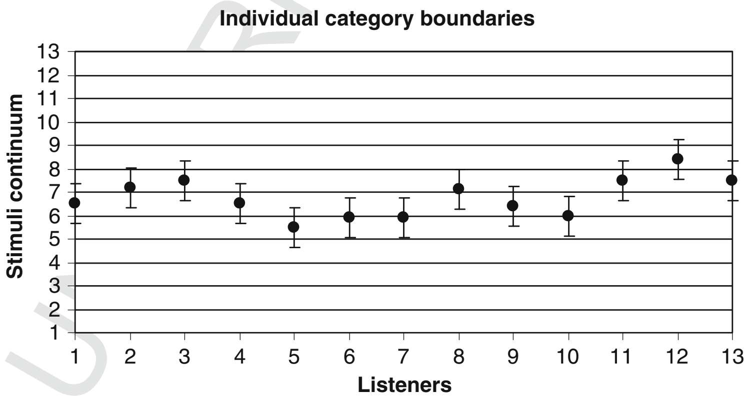


Fig. 5 Individual category boundaries for each of the listeners in the semantically motivated identification task

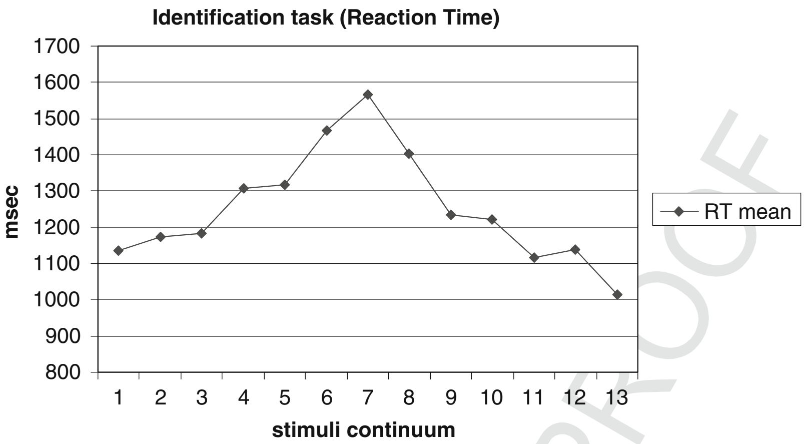


Fig. 6 Mean Reaction Time values in the semantically motivated identification task. A clear peak occurs at the category boundary determined in the labelling task responses, i.e. stimulus 7

4.4.2 Reaction Time

Reaction Time (RT) measurements provide further evidence of categorical interpretation. Mean values (Fig. 6) show a clear peak around the category boundary indicated by the identification judgements (stimulus 7), whereas RT mean values at the extremes of the phonetic continuum are shorter, i.e. listeners were less confident with more ambiguous stimuli, and more confident in labelling stimuli at the extremes of the continuum (i.e. less ambiguous).

These data are in line with previous results on similar perceptual experiments using RT measurements in categorical interpretation of intonation (Chen 2003; Vanrell Bosch 2006b; Falé and Hub Faria 2006; Gili Fivela 2008b; Borràs-Comes et al. 2010).

5 Discrimination task

5.1 Preparation and Presentation of Stimuli

In the discrimination task, 3 series of stimulus pairs were created, i.e. AB, BA and AA (false alarm set). The AB series (stimulus pairs 1-2, 2-3, 3-4, etc) and the BA series (2-1, 3-2, 4-3, etc) consisted of 12 stimulus pairs each, whereas the control AA series (1-1, 2-2, 3-3, etc.) consisted of 13 stimulus pairs. The inter-stimulus interval between each pair was 500 msec.

For each series, 3 repetitions were presented to listeners in a random order (plus an additional set of 37 stimulus pairs for training), for a total amount of 148 stimulus pairs. They were organised in blocks of 37 stimulus pairs, each group preceded by 2 warning tones and followed by 10 seconds of silence. Each stimulus pair was preceded by 1 warning tone and followed by 4 seconds silence for answering.

Informants were requested to judge whether the two utterances in each pair were the same or different, by pressing the appropriate button on the keyboard, i.e. 'U' for 'same' (Uguale) and 'D' for 'different' (Diverso), and – as in the identification task – this indication was kept available on the computer screen during the whole session.

Also in this case, subjects were asked to answer as quickly as possible, but in any case not before they had listened to the second of the two utterances in each pair. They were also warned they had a maximum of 4 seconds of time available for answering, and that it was not possible to skip any answer.

The discrimination task was performed right after the identification task, with few minutes break between the two tasks.

5.2 Results

5.2.1 Responses

The percentage of judgements of the pairs as 'different' in the discrimination task show a completely different trend – in terms of consistency in responses – as compared to the ones in the identification task. As is shown in Fig. 7, results indicate clearly that listeners were completely unreliable in their judgements. Although it might appear that there is a discrimination peak at the category boundary for AB hits, it cannot be considered significant as the percentage of agreement is only 38%, i.e. below the chance level.

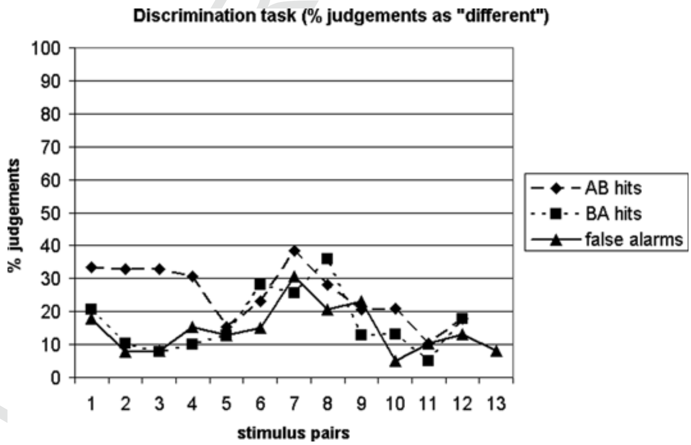


Fig. 7 Percentage of judgements as 'different' for AA (false alarms), AB hits and BA hits in the discrimination task. No clear patterns can be observed, listeners were unreliable in judgement

Unreliability in judgements is more evident if considering that response frequency for hits (AB and BA pairs) is not distinct from responses for false alarms (AA), clearly showing that listeners were unable to distinguish between them.

On the other hand, the fact that judgements on false alarms are not clearly below those on AB and BA hits cannot be ascribed to step size in the stimulus pairs, as 15 Hz cannot be considered as a small step size in perceptual terms.

In terms of unreliability, our results are more extreme compared to the above mentioned experiments for other languages (Ladd and Morton 1997; Remijsen and van Heuven 1999; Schneider and Lintfert 2003; Vanrell Bosch 2006a, 2006b), where subjects were at least more consistent in distinguishing between false alarms and hits, and where some indication of discrimination peak(s) can be observed.

5.2.2 Reaction Time

Even though we did not necessarily expect a discrimination peak in the responses, we had expected to obtain shorter RT in judging stimulus pairs around the category boundary as identified in the identification task by the listeners (i.e. less difficult to discriminate), and longer RT in judging stimulus pairs at the extreme of the phonetic continuum (i.e. more difficult to discriminate). Yet RT mean values (Fig. 8) do not show any valley, not even at the stimulus pair corresponding to the category boundary in the identification task. Moreover, it can be noted that RT values are very similar with both hits and false alarms, thus confirming unreliability and uncertainty of informants in their judgement.

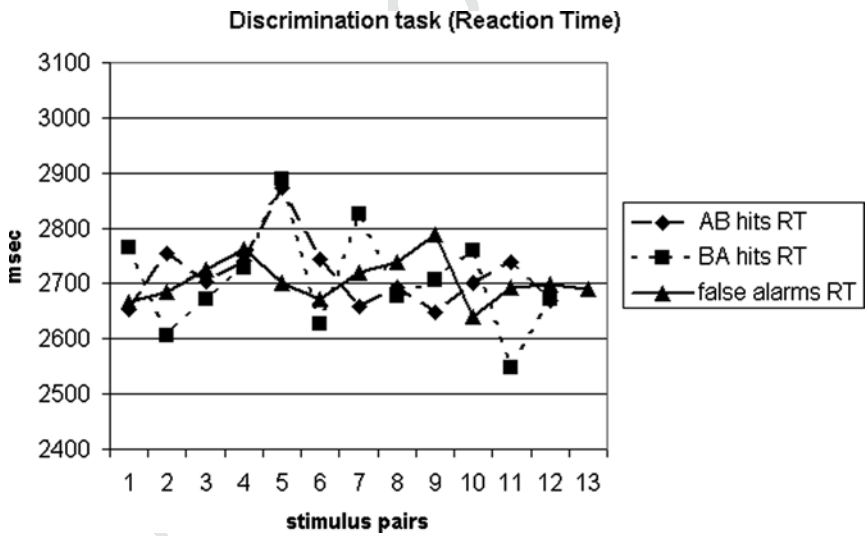


Fig. 8 Mean Reaction Times for AA (false alarms), AB hits and BA hits in the discrimination task. The expected valley at/around the category boundary determined in the identification task is not present, also similar results again for hits and false alarms, confirming unreliability and uncertainty in judgement

Our results are consistent with those obtained by Gili Fivela (2008a, 2008b) for Pisa Italian: in cases when no discrimination peak is found, RT values did not provide any evidence in terms of discrimination ability by the listeners (as in our experiment here).

6 Discussion and Conclusions

We have shown that Bari Italian listeners can reliably and consistently make a categorical interpretation of utterances as QUERY or OBJECT by listening to stimuli manipulated for peak height only. Such perceptual evidence is given by the typical S-shaped curve of response frequency, and also by the longest Reaction Time for judging stimuli corresponding to the category boundary in the phonetic continuum.

On the other hand, Bari Italian listeners appear to be completely unreliable and uncertain when asked to discriminate between pairs of stimuli, as results of the discrimination task demonstrate: not only is there no discrimination peak (this was expected, as in a number of previous similar experiments involving intonation no clear discrimination peak was found), but there is no clear trend in the Reaction Time measurements either.

Our results can be interpreted in the light of what has already been observed by Gerrit and Schouten (1998) for the perception of vowels. The authors conclude that vowel perception is only categorical when listeners are in the ‘phonetic mode’ i.e. when asked to classify speech stimuli in the labelling task, but they are unable to *discriminate* between pairs of stimuli, as in this case they are in the ‘psychoacoustic mode’, i.e. in a mode in which they do not access phonetic knowledge. They also claim that these results are not incompatible with ‘true’ categorical perception, as it can only occur when listeners are in the ‘phonetic mode’. In other words, they claim that categorical perception is obtained when subjects are accessing phonetic or linguistic knowledge.

Similarly, looking at our results from the identification and discrimination tasks together, we can see that Bari Italian listeners interpret peak height in a categorical way when required to make a judgement based on linguistic/pragmatic meaning. However, they show no evidence of categorical perception when asked to discriminate between stimuli. It appears therefore that they do not access their linguistic knowledge when performing the discrimination task.

Further experimental evidence that discrimination in intonation is mainly based on acoustic memory has been provided by studies like those described in Faulkner (1986) and Cummings et al. (2006). In these experiments, listeners were asked to discriminate between pairs of both speech and non-speech stimuli, after having performed a labelling task. Results in Faulkner (1986) show that informants indicate a discrimination peak (corresponding to the category boundary in the identification task) that was the same for both speech and non-speech stimuli, showing that the basis of the category boundary effect is psychoacoustic. In the perceptual experiment carried out by Cummings et al. (2006), discrimination peaks were inconsistent with category boundaries in the labelling task for both speech and non-speech stimuli, with a slightly better performance in terms of category boundary effect for non-speech stimuli. These results indicate that linguistic categories do not affect discrimination performance.

If discrimination is mainly based on psychoacoustic abilities, one factor influencing performance in discrimination could be listener specific competence (i.e. what training they have, e.g. whether they are phoneticians or musicians). In fact, Cummings et al. (2006) found a correlation between strong formal musical training and category boundary effects in a discrimination task. The background of our informants (no training at all), as opposed to that of the subjects in a number of experiments reported on in the literature (where subjects were students and staff of linguistics or phonetics institutes) might have contributed towards the poor performance of our subjects in the discrimination task. Indeed, if this is the case, it is all the more striking that they were able to perform so reliably in the identification task.

However, it is our view that the poor discrimination results are not simply a result of listener competence. If discriminating between pairs of stimuli mainly involves acoustic memory, there are a number of factors which might influence performance in tasks involving intonation. One of these factors might be that we are necessarily dealing with a pair of stimuli which have the length of one intonation phrase each. Even if each stimulus consists of a short phrase, it is difficult for the first of a pair to be retained in sensory memory once the second has been heard. At most, we would expect the final syllable or so to be retained, owing to an auditory recency effect (Conrad and Hull 1968, Crowder and Morton 1969; for comprehensive overviews see Neath 1998, Penney 1989). This is not the location of the distinction in our stimulus pairs. Instead it is the penultimate syllable. Issues relating to retention in sensory memory should thus be addressed in future studies.

In sum, despite the discrimination results, the reliable performance in the identification task of our naïve informants points to the necessity for the representation of [high peak] in the intonational phonology of this Italian variety.

Acknowledgments We would like to thank the audience at the third TIE conference in September 2008, for discussions of the experimental results. We are also grateful to Mariapaola D'Imperio and another three anonymous reviewers for helpful comments and suggestions on an earlier version of this paper. Many thanks also to Horst Lohnstein for discussion and advice on the semantics of questions, Ralf Rummer and Judith Schweppe for discussion of issues relating to sensory memory, and Mario Refice for help with statistical analysis. All errors are of course ours.

References

Anderson, Anne, Miles Bader, Ellen G. Bard, Elizabeth Boyle, Gwyneth Doherty, S. Garrod, Stephen Isard, Jacqueline Kowtko, Jan MacAllister, Cathy Sothillo, Henry Thompson, and Regina Weinert.1991. The HCRC Map Task Corpus. *Language and Speech* 34(4): 351–366.

Bartels, Christine. 1999. *The Intonation of English Statements and Questions*. New York: Garland.

Boersma, Paul, and David Weenink. 1999. Praat. A system for doing phonetics by computer. http://www.fon.hum.uva.nl/praat

Bolinger, Dwight. 1989. *Intonation and its uses*. Palo Alto: Stanford University Press.

Borràs-Comes, Joan, Maria del Mar Vanrell Bosch, and Pilar Prieto. 2010. The role of pitch range in establishing intonational contrast in Catalan. In *Proceedings of Speech Prosody 2010*, Chicago, 11-14 May 2010 (on CD-ROM).

Carletta, Jean, Amy Isard, Stephan Isard, Jacqueline Kowtko, Gwyneth Doherty-Sneddon, and Anne Anderson. 1997. The Reliability of a Dialogue Structure Coding Scheme. *Computational Linguistics* 23 (1): 13–32.

Chafe, William.1974. Language and Consciousness. *Language* 50: 111–133.

Chen, Aoju. 2003. Reaction Time as an indicator of discrete intonational contrast in English. In *Proceedings of Eurospeech 2003*, Geneva, 97–100.

Conrad, R., and A. J. Hull. 1968. Input modality and the serial position curve in short-term memory. *Psychonomic Science* 10: 135–136.

Crowder, R. G., and J. Morton. 1969. Precategorical acoustic storage (PAS). *Perception & Psychophysics* 5: 365–373.

Cruttenden Alan. 1986. *Intonation*. Cambridge: Cambridge University Press.

Cummings, Fred, Colin Doherty, and Laura Dilley. 2006. Phrase-final pitch discrimination in English. In *Proceedings of Speech Prosody 2006*, Dresden, 2-5 May 2006, (on CD-Rom).

Falé, Isabel, and Isabel Hub Faria. 2006. Categorical perception of intonational contrasts in European Portuguese. In *Proceedings of Speech Prosody 2006*, Dresden, 2-5 May 2006 (on CD-ROM).

Faulkner, Andrew. 1986. Categorical perception of speech intonation contour; the psychoacoustic basis of tonetic categories. *JASA* Suppl. 1, vol. 80, S126.

Gerrits, Ellen, and Bert Schouten. 1998. Categorical perception of vowels. In *Proceedings of the 5th International Conference of Spoken Language Processing*, Sydney, Australia, November 30-December 4, 1998, paper 0265.

Gili Fivela, Barbara. 2008a. Broad focus vs contrastive focus: is there categorical perception in Pisa Italian? In *Proceedings of Speech Prosody 2008*, Campinas (Brazil), 6-9 May 2008, 293–296.

Gili Fivela, Barbara. 2008b. *Intonation in Production and Perception: The Case of Pisa Italian.* Edizioni dell'Orso: Alessandria.

Grice, Martine, and Michelina Savino. 1995a. Intonation and communicative function in a regional variety of Italian. *PHONUS*, Institut fuer Phonetik/Phonologie, Universitaet des Saarlandes, vol. 1, 19–32.

Grice, Martine, and Michelina Savino. 1995b. Low tone versus 'sag' in Bari Italian intonation; a perceptual experiment. In *Proceedings of the XIII International Congress of Phonetic Sciences*, Stockholm 13-19 August 1995, vol. 4, 658–661.

Grice, Martine, and Michelina Savino. 1997. Can pitch accent type convey information status in yes-no questions?. In *Proceedings of the workshop sponsored by the Association of Computational Linguistics 'Concept-to-Speech Generation Systems'*, Madrid 14 July 1997, 29–38.

Grice, Martine, and Michelina Savino. 2003a. Question type and information structure in Italian. In *Proceedings of the International Workshop 'Prosodic Interfaces',* Nantes, 27-29 March 2003, 117–122.

Grice, Martine, and Michelina Savino. 2003b. Map Tasks in Italian: asking questions about given, accessible and new information. *Catalan Journal of Linguistics*, special issue on Romance Intonation (P. Prieto, editor), 2: 153–180.

Grice, Martine, and Michelina Savino. 2004. Information Structure and Questions – Evidence from Task-Oriented Dialogues in a Variety of Italian. In Peter Gilles and Joerg Peters, (eds.) *Regional Variation in Intonation*, 161–187. Tuebingen: Niemeyer.

Grice, Martine, Ralf Benzmueller, Michelina Savino, and Bistra Andreeva. 1995. The intonation of queries and checks across languages: data from Map Task dialogues. In *Proceedings of the XIII International Congress of Phonetic Sciences*, Stockholm 13-19 August 1995, vol. 3, 648–651.

Grice, Martine, Michelina Savino, and Mario Refice. 1997. The intonation of questions in Bari Italian: do speakers replicate their spontaneous speech when reading?, *PHONUS*, Institut fuer Phonetik/Phonologie, Universitaet des Saarlandes, vol.3, 1–7.

Grice, Martine, Mariapaola D'Imperio, Michelina Savino, and Cinzia Avesani. 2005. Strategies for intonation labelling across varieties of Italian. In Sun-Ah Jun, (ed.) *Prosodic Typology: The Phonology of Intonation and Phrasing*, 362–389. New York: Oxford University Press.

Hirschberg, Julia, and Gregory Ward. 1992. The influence of pitch range, duration, amplitude and spectral features on the interpretation of the rise-fall-rise intonation contour in English. *Journal of Phonetics* 20: 241–251.

Ladd, Robert D. 1994. Constraints on the gradient variability of pitch range, or, Pitch level 4 lives! In Patricia Keating, (ed.) *Phonological Structure and Phonetic Forms, Papers in Laboratory Phonology III* 43-63. Cambridge: Cambridge University Press.

Ladd, Robert D. 1996. *Intonational Phonology*. Cambridge: Cambridge University Press.

Ladd, Robert D., and Rachel Morton. 1997. The perception of intonational emphasis: continuous or categorical? *Journal of Phonetics* 25: 313–342.

Lieberman, Alvin M., Katherine S. Harris, Howard S. Hoffman, and Belver C. Griffith. 1957. The discrimination of speech sounds within and across phoneme boundaries. *Journal of Experimental psychology* 54(5): 358–368.

Neath, I. 1998. *Human memory: An Introduction to Research, Data, and Theory*. Pacific Grove, CA: Brooks/Cole.

Penney, C. G. 1989. Modality effects and the structure of short term verbal memory. *Memory & Cognition* 17: 398–422.

Pike, Kenneth L. 1945. *The Intonation of American English*. Ann Arbor: University of Michigan Press.

Pisoni, David B., and J. Tash. 1974. Reaction times to comparisons within and across phonetic categories. *Perception & Psychophysics* 15(2): 285–290.

Refice, Mario, Michelina Savino, and Martine Grice. 1997. A contribution to the estimation of naturalness in the intonation of Italian spontaneous speech. In *Proceedings of the V European Conference on Speech Communication and Technology (EUROSPEECH 97)*, Rhodos, 22-25 Sepember 1997, vol. 2, 783–786.

Remijsen, Bert, and Vincent van Heuven. 1999. Gradient and categorical pitch dimensions in Dutch: diagnostic test. In *Proceedings of the XIV International Conference of Phonetic Sciences*, San Francisco, 1865–1868.

Repp, Bruno. 1984. Categorical perception: issues, methods, findings. In Norman Lass (ed.) *Speech and language. Advances in Basic Research and Practice*, 243–335. New York: Academic Press.

Romero, Maribel. 2006. Biased Yes-No Questions: The Role of VERUM. *Sprache un Datenverarbeitung* 30: 9–24.

Romero, Maribel, and Chung-Hye Hann. 2004. On Negative Yes-No Questions. *Linguistics and Philosophy* 27: 609–658.

Savino, Michelina. 1997. *Il ruolo dell'intonazione nell'interazione comunicativa. Analisi strumentale delle domande polari in un corpus di dialoghi spontanei (varietà di Bari)*, unpublished PhD dissertation, Università di Bari/Politecnico di Bari, Italy.

Savino, Michelina. 2000. Descrizione autosegmentale-metrica di alcune tipologie intonative dell'italiano di Bari. In Elizabeth Burr, (ed). *Atti del VI Convegno Internazionale della SILFI (Società Internazionale di Linguistica e Filologia Italiana)*, Duisburg 28 June - 2 July 2000, *Tradizione & Innovazione. Linguistica e Filologia Italiana alle soglie del nuovo millennio*, 163–178, Firenze: Cesati (2006).

Savino, Michelina. 2001. Non-finality and Pre-finality in Bari Italian Intonation: a Preliminary Account. In *Proceedings of the VII European Conference on Speech Communication and Technology*, Aalborg 3-7 September 2001, vol. 2, 939–942.

Savino, Michelina. 2004. Intonational Cues to Discourse Structure in a Variety of Italian. In Peter Gilles and Joerg Peters, (eds.) *Regional Variation in Intonation*. 145–159. Tuebingen: Niemeyer.

Savino, Michelina, and Martine Grice. 2007. The role of pitch range in realising pragmatic contrasts – The case of two question types in Italian. In *Proceedings of the XVI International Conference of Phonetic Sciences*, Saarbruecken 6-10 August 2007, 1037–1040.

Schneider, Katrin, and Britta Lintfert. 2003. Categorical perception of boundary tones in German. In *Proceedings of the 15th International Conference of Phonetic Sciences*, Barcelona, 3-9 August 2003, 631–634.

Stirling, Lesley; Janet Fletcher, Ilana Mushin, and Roger Wales. 2001. Representational issues in annotation: using the Australian map task corpus to relate prosody and discourse structure. *Speech Communication* 33: 113–134.

Tench, Paul. 1996. *The Intonation Systems of English*. London: Cassell.

Vanrell Bosch, Maria del Mar. 2006a. A scaling contrast in Majorcan Catalan interrogatives. In *Proceedings of Speech Prosody 2006*, Dresden, 2-5 May 2006 (on CD-ROM).

Vanrell Bosch, Maria del Mar. 2006b. *The Phonological Role of Tonal Scaling in Majorcan Catalan Interrogatives*, MA thesis, UAB, Barcelona, Spain.

Ward, Gregory, and Julia Hirschberg. 1988. Intonation and Propositional Attitude: The Pragmatics of L* + H L H%. In *Proceedings of the Fifth Eastern States Conference on Linguistics*, 512–522.

From Tones to Tunes: Effects of the f_0 Prenuclear Region in the Perception of Neapolitan Statements and Questions

Caterina Petrone and Mariapaola D'Imperio

1 Introduction

According to the Autosegmental-Metrical (AM) Theory (Pierrehumbert 1980; Pierrehumbert and Beckman 1988; Ladd 2008), the intonation contour can be represented as a sequence of pitch accents and edge tones, which are generated by a finite-state grammar (Pierrehumbert 1980). Nuclear and prenuclear accents are selected from one and the same inventory, and their distinction is merely positional, the nuclear accent being the last accent in the intermediate phrase. From a computational point of view, Pierrehumbert's grammar is also 'non deterministic', since it does not give any information about the transitional probabilities between one state and the following one. In other terms, pitch accents and edge tones can occur in any combination. Such a free compositionality is due to the fact that the meaning of the intonational contour is the sum of the independent contribution of each of its tonal morphemes (Pierrehumbert and Hirschberg 1990). The domain of interpretation of each morpheme corresponds to its phonological domain. So, for example, in American English, H* accents signal that the accented item is new, while L* accents are employed for giving salience to items which the Speaker believes to be already part of the Hearer's beliefs. On the contrary, edge tones are employed for highlighting the relationship between the propositional content of the current intermediate/intonational phrase and that of previous or following ones.

Though, according to Pierrehumbert and Hirschberg (1990), as well as to other theories of meaning compositionality (Gussenhoven 1984; Bartels 1999; Steedman 2003; Marandin et al. 2004, *inter alia*), all tonal morphemes of a tune should contribute to its meaning, it is implicitly maintained that the nuclear configuration (i.e., the intonation region including the nuclear accent, the phrase accent and the boundary tone) is the semantic 'heart' of tunes. Specifically, Pierrehumbert and Hirschberg (1990) describe the combination of the

C. Petrone (✉)
Zentrum für Allgemeine Sprachwissenschaft, Berlin, Germany
e-mail: petrone@zas.gwz-berlin.de

S. Frota et al. (eds.), *Prosodic Categories: Production, Perception and Comprehension*, Studies in Natural Language and Linguistic Theory 82,
DOI 10.1007/978-94-007-0137-3_9, © Springer Science+Business Media B.V. 2011

Table 1 Pitch accents functions and their use in American English according to Pierrehumbert and Hirschberg's theory of intonational meaning (taken from Pierrehumbert and Hirschberg 1990).

Accent	Meaning	Typical use	Examples
H*	The item must be treated as 'new' in the discourse	Declarative	You turkey H* L- L% You deliberately deleted my files H* H* H* L- L%
L*	The item is not to be instantiated in the open expression to be added to Hearer's mutual beliefs	Yes/no question	Do prunes have feet? L* L* H- H%

nuclear H* accent with L- L% as the typical tune for declarative sentences, while the combination of the nuclear L* accent with H- and H% would be typical for yes/no questions:[1] However, no particular attention has been given to define the exact contribution of the f_0 prenuclear region to tune meaning. For instance, in utterances containing more than one pitch accent, the prenuclear accents are mere copies of the nuclear one (Table 1). The authors also admit the difficulty in analyzing utterances with mixed accent types in the prenuclear and nuclear contours and defer the question to future research.

In Italian, the positional definition of the nuclear accent has been revised as the 'rightmost fully fledged pitch accent in the focussed constituent' (Grice et al. 2005), since it can be followed by postnuclear accents within the same intermediate phrase. Moreover, in the Neapolitan variety of Italian, the nuclear accent also 'encapsulates' the distinction between questions and statements, the boundary tone being in both the cases L- L%. Specifically, the contrast between late (L* + H) and early (L + H*) nuclear accent alignment is employed to signal yes/no questions vs. narrow focus statements. This difference is also a very robust cue for intonation identification (D'Imperio and House 1997; D'Imperio 2000; Petrone 2008).

Recently, we also found an acoustic difference between yes/no questions and narrow focus statements in the region spanning from the prenuclear (LH*) accent to the nuclear rise (Petrone and D'Imperio 2008; D'Imperio and Petrone 2008; Petrone 2008). According to the AM theory, if no intermediate tone is present, a linear interpolation is expected between the H tone of the prenuclear rise and the leading L tone of the nuclear rise. However, a different picture is offered by Neapolitan Italian, as shown in Fig. 1. Specifically, in statements (Fig. 1, upper), the f_0 rapidly falls from the prenuclear H to the region immediately after the end of the first prosodic word, with a low turning point followed by a f_0 plateau which continues until the beginning of the nuclear rise. On the other hand, in questions (Fig. 1, lower), the f_0 fall after the prenuclear H is much

[1]Note that, in such compositional view, the relationship between tunes and speech acts (like 'assertion' or 'question') is not one-to-one, so that the same tune can be associated to different speech acts, and *vice versa*. For example, the H* L- L% pattern might occasionally be employed with *wh*- questions in American English.

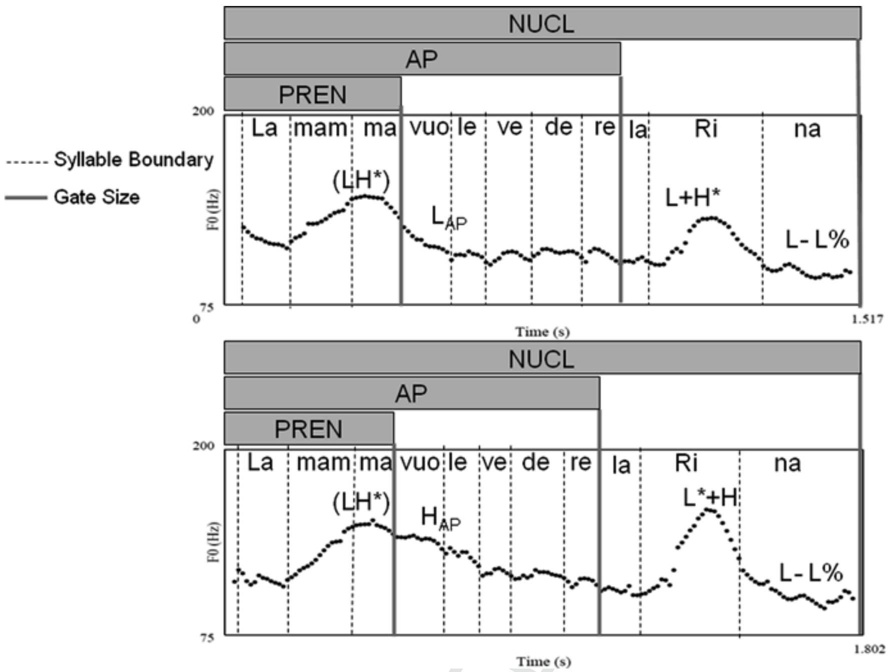


Fig. 1 Schematized representation of the stimulus manipulation (three conditions: PREN, AP, NUCL) for the sentence *La mamma vuole vedere la Rina* ('(The) mom wants to see (the) Rina'), uttered as a narrow focus statement with late focus (upper picture), and as a yes/no question (lower picture)

shallower, so that the f_0 contour in the immediate postaccentual region results into a convex downward parabola. After this region, the slope becomes steeper in order to attend the low values for the L* of the L* + H nuclear accent. Our acoustic results reveal that, though such a difference in the shape and slope is systematic, the inflection point of the two curves occurs at similar temporal locations (i.e., around the right edge of the prosodic word), while being melodically lower in statements than in questions (Petrone and D'Imperio 2008; D'Imperio and Petrone 2008). This lead us to the hypothesis that the difference in the interaccentual slope might be due to the insertion of a tone, differently specified in the two intonation modalities. This tone would mark the end of the Accentual Phrase (AP), in both questions and statements, though having a different tonal specification (H_{AP} for questions and L_{AP} for statements).[2]

[2]Another possible hypothesis is that the tone following the prenuclear peak is part of the prenuclear accent. This would lead us to reanalyse the prenuclear accent as tritonal, with a different phonological specification for statements (LH*L) and questions (LH*H or even LH*!H). Apart from theory-internal type of arguments against such a proposal (see Grice 1995), preliminary studies have also found that, in Neapolitan, this tone is consistently aligned

Hence, a question raised by our data is: why would the AP tonal specification be different in questions and statements? In this paper, we argue that this difference would help Neapolitan listeners to recover the contrast between yes/no questions and statements even when nuclear contour information is not available. This hypothesis stems from two observations. First, in Italian, the contrast between questions and statements is signalled solely by prosodic means, while no morpho-syntactic cues are generally employed to distinguish the two modalities. Moreover, in Neapolitan Italian, the nuclear alignment contrast between yes/no questions and narrow focus statements is very subtle, the nuclear peaks being, on average, only 45 ms later in the question modality (D'Imperio 2000). Therefore, we might hypothesize that the use of a different tone (H_{AP} vs. L_{AP}) in the prenuclear contour would enhance the phonological contrast between the two modalities.

To verify this, we carried out two perception experiments, in which auditory speech stimuli were gated at different locations of the sentence. The first experiment was aimed at determining whether Neapolitan listeners are able to distinguish questions and statements in absence of the nuclear accent. Natural stimuli were employed and listeners' responses were elicited by means of an identification task. The second experiment was a semantic differential task, aimed at investigating the semantic properties of the AP tone, and its contribution to the perception of the intonation contrast.

2 Experiment I

2.1 Methodology

2.1.1 Stimuli Preparation

The stimulus set was created from three natural utterances, selected from a corpus of read sentences (for details, see Petrone and D'Imperio 2008; Petrone 2008), such as *La mamma vuole vedere la X* 'Mom would like to see X'. Utterance stimuli were composed of: (1) the utterance-initial noun phrase *La mamma* ('the mom'), in which the stressed syllable *mam-* carried a prenuclear accent in both intonation modalities; the unaccented verbal phrase *vuole vedere* ('wants to see'), in which the AP tone was realized on the syllable *vuo-* of *vuole*; and a paroxyton proper name in utterance-final position (*la Dina* / *la Rina* / *la Bina*), bearing the nuclear rise. Each sentence was uttered by a native speaker of Neapolitan Italian (OM) both as yes/no question and a narrow focus statement with late (narrow)

---

around the end of the prosodic word, independently of the prenuclear rise temporal location (Petrone 2008), thus suggesting that we are dealing with an edge tone. Moreover, D'Imperio and Petrone (2008) and Petrone (2008) found that this tone is not accompanied by the percept of an intermediate phrase break nor by a degree of final lengthening comparable to that at the end of an intermediate phrase, thus suggesting that it would mark the right boundary of a smaller prosodic constituent, i.e., the AP.

focus placement. As a consequence, the two intonation modalities differed in the phonological specification of the prenuclear AP tone (H_{AP} vs. L_{AP}) as well as in nuclear accent category (L + H* vs. L* + H). The prenuclear accent was rising in both modalities and, following work by Prieto et al. (2005), we label it as (LH*).[3] In order to isolate the contribution of the prenuclear f_0 region to intonation identification, the tonal composition of each stimulus was manipulated. Specifically, each stimulus was gated at two different locations in the sentence, i.e., at the end of *La mamma* and at the end of *vuole*. The first group of stimuli contained only the prenuclear rise and a portion of the f_0 transition from the prenuclear H to the following AP tone (PREN condition). The second group contained both the prenuclear accent and the entire AP tone (AP condition). The entire utterance was also included as a control, since only in these stimuli the nuclear accent configuration was available to the listeners (NUCL condition). The three experimental conditions are shown in Fig. 1.[4]

Stimulus duration was also slightly different across intonation modality. Specifically, stimuli created from question (Q) bases were shorter than those created from statement (S) bases, in both conditions PREN (mean value across repetitions: Q = 380 ms; S = 393 ms.) and AP (Q = 573 ms.; S = 592 ms). However, in the NUCL condition, utterances tended to be globally longer in the question (1.486 ms) than in the statement (1.448 ms) bases series. This is in agreement with data on Neapolitan by Petrone (2008), who found that word duration is shorter in questions when the word is associated with a prenuclear accent, while they are shorter in statements when associated with a nuclear accent.

2.1.2 Task and Analysis Procedure

The 18 stimuli (3 sentences X 2 intonation modalities X 3 tonal gates) were played directly from a laptop computer by means of PERCEVAL, a software for performing computerized auditory and visual perception experiments (André et al. 2007) developed at the *Laboratoire Parole et Langage* (Aix-en-Provence, France). All stimuli were presented binaurally through professional headphones (Senneheiser HD 497) in a quiet room. Two buttons *Domanda* ('question') and *Affermazione* ('statement') were visualized on the computer screen in the same order to avoid uncertainty in responses due to order shifts.

The stimulus group was played 5 times in 3 separate randomized blocks, containing respectively stimuli from the PREN, AP and NUCL conditions.

[3]We label the prenuclear rise as (LH*) to distinguish it from the nuclear L + H* accent of narrow focus statements. Specifically, while the peak in the L + H* accent seems to have a secondary association with the first mora in the accented syllable, the peak in the (LH*) does not have any secondary association with segmental anchors in the metrical structure (Prieto et al. 2005).

[4]Though our procedure is reminiscent of the gating paradigm from the segmental literature (Grosjean 1980; Lahiri and Marslen-Wilson 1991, *inter alia*), the choice of cutting the stimuli at the end of the word (instead of at sub-word locations) allowed us to obtain more natural stimuli.

Repetitions of the same stimulus within each block were also randomly played, with order of presentation varying across listeners and blocks. This helped avoiding possible order of presentation effects (Savino and Grice 2007; 2008). The start of each block presentation was preceded by a visual message on the screen. Moreover, the stimuli were all separated by a four-second pause; a sixty-second pause followed the end of each block. For the specific instructions, listeners were told that they were going to listen to some sentences (for the NUCL condition block), or to just a fragment of a sentence, as if the speaker was suddenly interrupted while formulating it (for the PREN and AP condition blocks).

Subjects performed a two-alternative forced choice task, in which they were asked to label each stimulus as either a question or a statement. They indicated their choice by clicking the right arrow on the keyboard for questions and the left one for statements. The task was preceded by a trial session, in which listeners had to identify ten sentences containing a prenuclear and a nuclear accent, uttered as either questions or statements. These sentences were not gated, since our aim was to familiarize listeners with the identification task. Practice trials were randomly selected from another Neapolitan corpus of read speech (Petrone and Ladd 2007) and were also randomly presented. The choice of presenting stimuli from the three experimental conditions (PREN, AP and NUCL) in three subsequent blocks was adopted to avoid possible 'learning' effects during the identification task. Specifically, since, in the AP block, stimuli contained both the prenuclear rise (already present in stimuli in the PREN condition) and the following AP tone, we assumed that judgments for the AP block would not be influenced by judgments of the PREN block. The experiment lasted less than 10 minutes.

2.1.3 Participants

Nine listeners participated in the experiment, two females and seven males. The listeners, who were not paid for their participation, were all brought up in Naples and spoke standard Italian with a Neapolitan accent. All the participants were between 20 and 30 years old. Two of them were students in linguistics.

2.2 Results

If question/statement identification depends solely on the availability of nuclear accent information, listeners should be capable to identify stimuli only in the NUCL condition, while scores would be at chance level in the PREN and AP conditions. However, Fig. 2 shows a different picture. In this histogram, percentages of 'question' responses for all subjects are shown across the three steps of tonal gate manipulation.

When listening to stimuli in the PREN condition, Neapolitans were already able to distinguish questions from statements. Specifically, mean 'question' score for question base stimuli was already above chance (67%), while it was

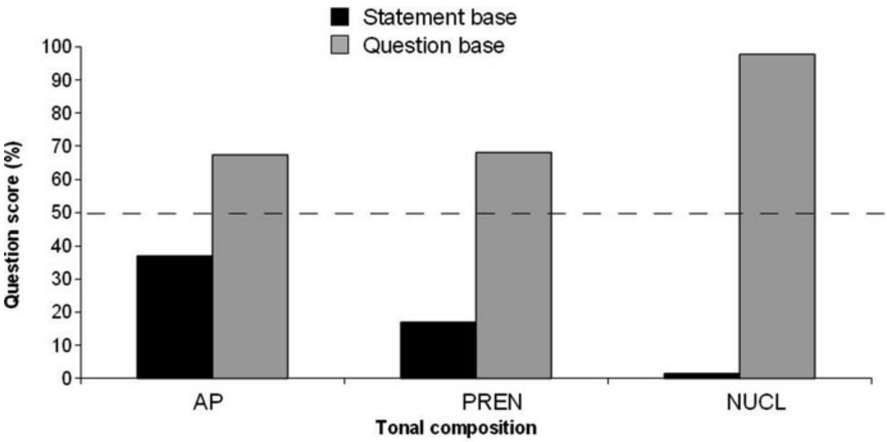


Fig. 2 Identification score for statement (black) and question (grey) base stimuli plotted separately for PREN, AP and NUCL conditions. Results are pooled for all listeners. The dotted line indicates chance level (50%)

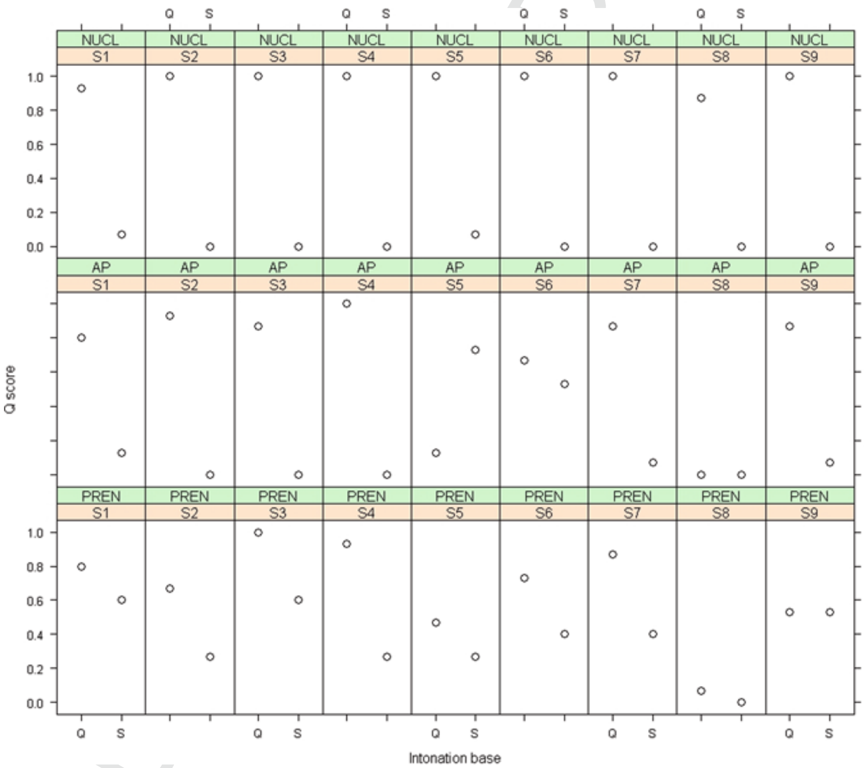
around 37% for statement base stimuli. In the AP condition, question scores decreased for statement base stimuli (20%), suggesting that the presence of the L_{AP} tone might have played a role in identification. However, question base stimuli scored similarly to the PREN condition (68%). The graph also shows that question judgments for question base stimuli only increased in the NUCL condition in a significant way. This indicates that the nuclear accent information is still important for question identification, though not necessary, since scores for question identification were well above chance already for segments lacking the nuclear accent. The presence of the nuclear accent also contributed to a drastic lowering of question responses for statement base stimuli (10%).

The statistical analysis, performed by means of the R-environment (R Development Core Team 2008), included a series of linear mixed models (Pinheiro and Bates 2000), in which (modality) Base Type (Q/S) and Gate Size (PREN/AP/NUCL) were our two fixed factors, whereas Listeners was the random factor. Since Gate Size had three levels, the effects of Base Type for the PREN and AP condition on 'question' score responses were evaluated by running two linear mixed models: one with PREN and the other with AP as reference level. Because of the complexity of the multiple analyses performed on the dataset, we used an alpha of $pMCMC < .01$.[5]

[5] In statistics, it is still unclear how to calculate the number of degrees of freedom in regression models including random factors. In mixed models, a valid alternative to "standard" *p*-values is to calculate the *p*-value from a MONTE CARLO sampling by Markov chain ($pMCMC$ = Monte Carlo Markov Chain; see Baayen 2008). Such values, automatically calculated by the *lme4* R package, are reported here to evaluate the statistical significance of the fixed factors in our models.

Our results showed a significant effect of Base Type already in both the PREN [t = −6.8, *pMCMC* < .01] and AP condition [t = −11.7, *pMCMC* < .01], thus confirming the existence of cues for intonation modality already in the region containing the prenuclear rise and the AP tone. Moreover, scores for question responses were significantly lower in the AP than in the PREN condition, but only for statement base stimuli [t = 3.35, *pMCMC* < .01]. On the other hand, no difference between the two conditions was found for question base stimuli [t = −0.17, *pMCMC* = .86]. Moreover, scores for question responses was lower in the AP condition relative to the NUCL one in statement base stimuli [t = −7.31, *pMCMC* < .01], while it significantly increased in questions base stimuli [t = 6.78, *pMCMC* < .01].

We also checked whether judgments across listeners were consistent. In Fig. 3, mean question scores (y-axis) are plotted by listener for the two intonation bases (x-axis), separately for the NUCL, AP and PREN conditions.While listeners were all able to well identify stimuli in the NUCL condition, some discrepancies can be noted in the PREN and the AP conditions. In the PREN condition, mean



This figure will be printed in b/w

Fig. 3 Mean question identification score by subject for question and statement base stimuli plotted separately for NUCL, PREN and AP conditions. The question score is bounded between 0 and 1, which is equal to 0% and 100% question score, respectively question score for question base stimuli was already around 80% for 4 listeners (S1, S3, S4, S7) out of 9, while it was at 50% (chance level) or even below it for 3 listeners (S5, S8, S9). In the same condition, mean question score for the S base stimuli was already around 20% for 4 listeners (S2, S4, S5, S8) and at 50% or below for 3 of them (S6, S7, S9). In the AP condition, 80% question score for the question base stimuli is reached by two listeners (S2 and S9). Higher consistency among subjects was obtained for statement base stimuli, in which mean question score was 20% for 8 listeners out of 9.

2.3 Discussion

The results of Experiment I suggest that Neapolitan listeners might identify yes/no questions and narrow focus statements even when the main cue for such a contrast (i.e., the alignment of the nuclear accent) is not available. First, utterance fragments solely containing the prenuclear accent successfully cue the intonation contrast, independently of base stimulus type (question/statement). This cannot be due to a difference in the tonal specification of the prenuclear accent, as it is rising (LH*) in both question and statement bases. Since in this task we employed natural stimuli, many cues might have been employed by listeners for the intonation contrast identification. For example, phonetic differences in the utterance-initial f_0 value between questions and statements have been reported in German (Brinckmann and Benzmueller 1999) and Dutch (van Heuven and Haan 2000). The impact of cues other than f_0 on intonation identification has been also attested in the literature (Pierrehumbert and Steele 1987; Haan 2001; van Heuven and van Zanten 2005, *inter alia*) and an effect of base stimulus on the question/statement contrast identification has already been found by D'Imperio (2000) and D'Imperio, Cangemi and Brunetti (2008) for Neapolitan Italian. However, the search of such cues is beyond the scope of this paper, and it will not be explored further here.

An important result concerns the AP condition, in which utterance fragments contained both the prenuclear rise and the AP tone (specified as L_{AP} in statements and H_{AP} in questions). If, for these stimuli, the perception of questions and statements depended only on phonetic differences in the realisation of the prenuclear rise or on factors other than f_0, question response scores would have been similar to those obtained for the PREN condition. On the contrary, in stimuli created from statement bases, question identification was significantly lower than in the PREN one, thus suggesting that the additional presence of the L_{AP} caused a shift towards statement interpretation. The results for question base stimuli are more puzzling. For these stimuli, question responses were always significantly above the chance level, meaning that listeners successfully completed the identification task in all cases. However, this result cannot be explained by the additional presence of the H_{AP}, since scores were similar for PREN and AP conditions. A possible hypothesis to explain such a discrepancy between the two intonation modalities is that the steep postaccentual fall is a clear cue for the statement identification, whereas, when such fall is absent or much shallower, stimuli would score more 'question' responses as a default choice.

Moreover, we could also ask whether our results might have been influenced by the experimental paradigm employed (i.e., an identification task for gated stimuli) and whether the identification task is appropriate to capture the meaning conveyed by the scaling variations in the prenuclear fall between questions and statements. The applicability of the categorical perception (CP) paradigm to intonational contrasts is indeed a controversial issue in the literature (see Gussenhoven 2004 for a short review). For instance, a common result is that, when a category boundary is found in the identification task, a corresponding discrimination peak is not found in the subsequent discrimination task, since in the second case perceptual judgment is mainly based on acoustic memory (Niebuhr and Kohler 2004; Scheider 2006, *inter alia*). A good alternative which has been proposed is that of measuring Reaction Times (RT) in the identification task (Chen 2003; Niebuhr 2007). Since we used gated stimuli, RTs were not employed in our experiment. In fact, in a similar study (Petrone 2010), we already found that, in an identification task, RT differences across gated stimuli were never significant, probably due to task difficulty.

We might also think that in Neapolitan the CP paradigm cannot account for the semantic space covered by the AP tone, and thus for its exact contribution to the perception of the contrast between questions and statements. An alternative method is the semantic differential task. Though this task is less widespread, it seems quite 'promising' for studying the association between intonational form and function. In fact, different from a forced-choice task, it is a multidimensional rating task, in which listeners rate their judgments on various semantic scales, which can be both linguistic and paralinguistic. Scales can be analogic or even continuous. Such a task has been employed in the Experiment II.

Finally, since in this experiment we used natural stimuli, we still do not know to what extent the scaling difference between L_{AP} and H_{AP} affect the perception of the intonation contrast. AP scaling was thus manipulated in Experiment II.

3 Experiment II

In this experiment, we explored of a possible effect of scaling differences between L_{AP} and H_{AP} by means of a semantic differential task (Uldall 1960; Grabe et al. 1998; Dombrowski 2003; Rathcke and Harrington (2010), *inter alia*). Specifically, five semantic scales were selected on the basis of hypotheses about the linguistic and paralinguistic properties of the two AP tones. This choice was due to the fact that we still do not know which kind of meaning is carried by the AP tone, i.e., whether it contributes in building the linguistic contrast between questions and statements or whether it carries a paralinguistic or attitudinal meaning. The existence of affective morphemes has already been found by

Grabe et al. (1998) for Dutch. Specifically, they found that in this language the % H tone (or 'high prehead', as opposed to a %L tone or 'low prehead') carries the meaning of 'sociability' and 'politeness', but it does not carry any linguistic meaning, i.e., relative to the pragmatic content of the message (but see Gussenhoven 2004 for a different interpretation of Dutch affective morphemes). If, in Neapolitan, the AP tone is employed to reinforce the contrast between questions and statements, we expect that its presence will influence listeners' judgments on the linguistic scales. On the contrary, if it has a mere affective value, it will influence solely responses on the paralinguistic scales.

3.1 Methods

3.1.1 Stimuli Preparation

The corpus was composed of stimuli which were resynthesized from one statement base utterance (speaker OM), *La mamma vuole vedere la Dina*. This sentence, already included in the stimulus set of Experiment I, was produced with two accents, a prenuclear (LH*) accent on the syllable *mam-* of *mamma* and a nuclear L + H* accent on *Di-* of *Dina*; the interaccentual region was characterized by the insertion of a L_{AP} tone (Fig. 1, upper panel).

The construction of the stimuli was based on the idea that the main cue for the question/statement distinction in the prenuclear f_0 region is the difference in tonal scaling between the L_{AP} and the H_{AP} tones. First, the utterance was cut at the onset of the nuclear syllable to prevent listeners from exploiting the alignment information carried by the nuclear accent in identifying questions vs. statements. Therefore, a linear stylization of the pitch contour was carried out (see Fig. 4), in which five points were interpolated: one point at the utterance beginning; two points at the beginning and end of the prenuclear f_0 rise; one point at the L_{AP} temporal location and one point at the fragment end (corresponding to the temporal position of the nuclear L). Pitch values for the utterance beginning, the prenuclear accent and the nuclear L tone were intermediate between those of a typical question and a typical statement for that speaker. Specifically, the f_0 values at the start of the utterance and at the start of the prenuclear rise were fixed at 114 Hz, the prenuclear f_0 peak was fixed at 157 Hz, the following L_{AP} at 120 Hz and the nuclear L at 103 Hz. The alignment of the prenuclear L and H target as well as that of the nuclear L were kept the same as in the natural productions, since in a previous study (Petrone 2008) we found that, for speaker OM, the temporal location of these targets corresponds with the stressed syllable onset independently of intonation modality. Once the stylization was applied, tonal scaling was modified.

Specifically, in order to verify whether intonation identification is affected by the contrast between L_{AP} and H_{AP}, a continuum in the tonal scaling domain was designed to cover the two phonological categories. Hence, the continuum was created by progressively raising the L_{AP} height and by connecting it to the

Table 2 Steps for the tonal scaling manipulation of the AP tone for the phonetic continuum in the semantic differential task

Scaling steps	H1	H2	H3	H4	H5	H6	H7	H8	H9	H10
f_0 values (Hz)	120	123.5	127.1	130.9	134.7	138.6	142.7	146.9	151.1	155.6

preceding prenuclear H and to the following nuclear L by means of straight lines. Specifically, we raised the L_{AP} height, until a value which was similar to that of the prenuclear peak, in ten 0.5 semitones steps (H1-H10). Following t'Hart (1981), we assumed that only tonal falls spanning more than three semitones can be clearly discriminated by listeners (see Rathcke and Harrington, 2010 for a similar approach to tonal scaling manipulation). As a consequence, we expected the first five steps of the continuum (H1-H5) to correspond to the values for L_{AP} and the last five (H6-H10) to H_{AP}. The ten stimuli were then resynthesized through PSOLA (Moulines and Charpentier 1990), in order for them to sound more natural. Scaling values are presented in Table 2.

The stimuli were cut at the same temporal locations as in Experiment I: at the end of *La mamma* (PREN condition) and at the end of *La mamma vuole* (AP condition). Consequently, stimuli in the PREN condition contained only the prenuclear rise and a short portion of the following f_0 fall, while stimuli in the AP condition presented the entire f_0 fall, including the AP tone. If the contrast between L_{AP} vs. H_{AP} plays a role in the identification of questions vs. statements, listeners'judgments will be affected by the scaling manipulation only in the AP condition, while they will be around the chance level in stimuli of PREN condition (where the AP tone was not present). Figures 4 and 5 show a

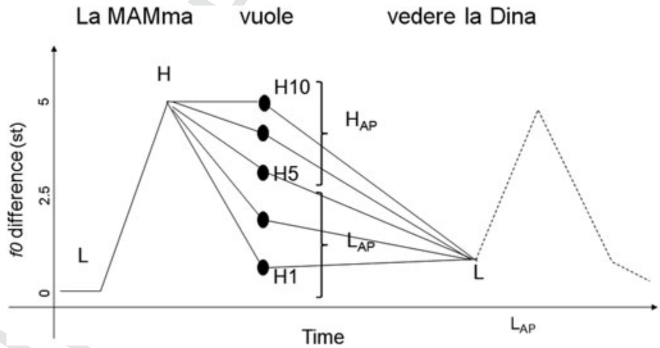


Fig. 4 Representation of the tonal scaling continuum created by raising the *f0* height of the L_{AP} tone. The association between the scaling steps and the AP tone is indicated to the right. For ease of schematization, the scaling manipulation is represented for only 5 steps of the continuum. The dotted line represents the nuclear accent, which was omitted from the gated stimuli

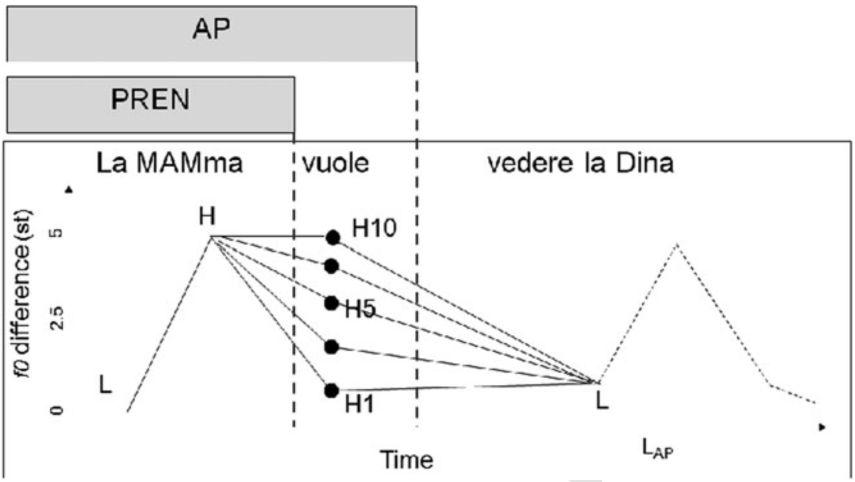


Fig. 5 Schematized representation of the two gate sizes (PREN, AP)

schematized representation of the scaling manipulation of L_{AP} and of the gating, respectively.

Finally, two control sentences were added to the stimulus set, in which the values at the extreme edges of the continuum were combined with the rise-fall nuclear configuration of statements and questions. Specifically, the temporal alignment and scaling of the peak in the LH nuclear rise as well as of the following L- phrase accent corresponded to mean values for that speaker (Fig. 6). These two stimuli were also resynthesized through PSOLA. The inclusion of such stimuli was aimed at controlling the possible semantic contribution of L_{AP} vs.

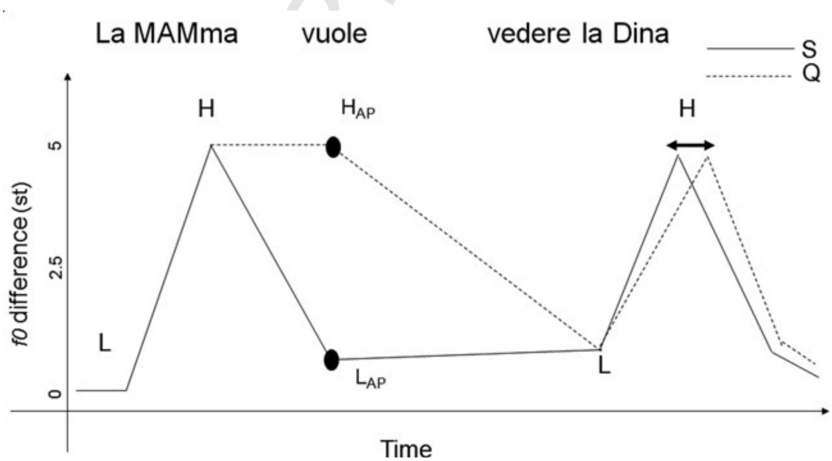


Fig. 6 Schematized representation of the two control sentences. The continuous line indicates the statement stimulus, while the dashed line indicates the question stimulus

H_{AP}. It is important to notice that, since in this experiment we employed only one base stimulus (i.e., a statement), we excluded that listeners would employ variability in prosodic cues other than f0 for intonation identification in the AP and PREN conditions as well as in the two control sentences.

This corpus, composed of a total of 22 stimuli (10 scaling steps * 2 gates + 2 control sentences), constituted the basis for our perception experiment.

3.1.2 Semantic Scales

Five semantic scales were chosen on the basis of *a priori* hypotheses about the semantic properties (linguistic and paralinguistic) of the L_{AP} and H_{AP} tones and in particular about their contribution in building the meaning contrast between questions and statements.

The first scale ('commitment') is based on a linguistic hypothesis about the role of final low vs. high tones in discourse. As far as we know, there are still no studies on the contribution of tonal morphemes in discourse interpretation in Italian. Therefore, this scale is inspired from the notion of 'commitment' already employed in Pierrehumbert and Hirschberg (1990) for American English as well as recent proposals by Marandin et al. (2004) and Marandin (2006) for French, in which specific tonal morphemes indicate how, according to the speaker's beliefs, the listener should/would interpret the message. Specifically, following Marandin et al. (2004), when the speaker employs a low tone, he signals that the prepositional content of his message will be accepted by the listener, independently of whether his beliefs are compatible with those of the listener or not. Therefore, the speaker also thinks that the discourse will be continued by the listeners as a function of such a message. When the speaker employs an edge-final high tone, on the contrary, he signals that his beliefs are *not* compatible with those of the listener, and that he commits himself in revising the propositional content of his message. The labels chosen to indicate such a contrast are *contestabile* / *incontestabile* ('contestable'/ 'incontestable').

The second scale ('potency') is based on the linguistic (or 'informational', Gussenhoven 2002) interpretation of the Frequency Code (Ohala 1983; Gussenhoven 2002). Low vs. high tones indicate that the speaker is certain vs. uncertain about the content of his message, and so that he is 'asserting' vs. 'questioning'. The labels chosen for this scale were *insicuro* / *sicuro* ('uncertain'/ 'certain'). The third ('activity'), fourth ('evaluation') and fifth ('submission') scales are inspired by the paralinguistic (or 'attitudinal', Gussenhoven 2002) interpretation of the Frequency Code, according to which low tones convey speaker's detachment, hostility and dominance, whereas high tones convey speaker's emotional involvement, sociability and submission. The labels chosen for these scales are *distaccato* / *coinvolto* ('detached' / 'involved'), *amichevole* / *ostile* ('friendly' / 'hostile') and *sottomesso* / *autoritario* ('submissive' / 'authoritative').

Given the meaning distinctions associated with the linguistic and paralinguistic interpretation of the Frequency Code, we expected the left pole of each scale to be associated by listeners to H_{AP} and the right one to $L_{AP.}$

3.1.3 Experimental Procedure

The stimuli were presented binaurally via Sennheiser HD 497 headphones in a silent room in the house of one of the speakers. The set of stimuli was played through a laptop by means of the PERCEVAL software package. The subjects had to listen to the stimuli for the PREN and AP condition presented in the same randomized block. At the end of the session, they listened to the control sentences. This presentation order was intended to avoid 'learning' effects, which could have biased listeners' responses for the stimulus fragments.

Each stimulus was heard once for each scale. The scales were visualized consecutively on the laptop screen. Listeners were asked to rank the stimuli between −3 and +3, i.e., between the extremes of each scale. They were told to assign these values as follows: '0' as 'neutral', '+/– 1' as 'slightly', '+/– 2' as 'quite' and '+/– 3' as 'very'. To reinforce the auditory impression of the stimuli, they were played twice consecutively with a two-second pause between the two repetitions. Listeners could answer only after the second repetition.

At the beginning of the experiment, subjects were told that they were going to listen to sentences or fragments of sentences, as if they suddenly arrived into a room in which a person was talking to another one. In this scenario, the speaker interrupts his sentence as soon as he hears the noise of the door opening. Subjects had therefore to judge how the speaker sounds like by clicking one of the seven buttons corresponding to the seven points in the semantic scales visualized on the laptop screen. The instructions were also visualized on the screen at the beginning of the experimental session.

To familiarize the listeners with the semantic scales and with the experimental procedure, the stimulus set was preceded by 10 practice trials selected among those in the PREN and AP condition. As in the main experiment, each stimulus was associated with a semantic scale. At the end of the training session, listeners were allowed to ask questions about the scales as well as the procedure. Also, at the end of the main experiment we asked subjects to give us their opinion about the difficulty of the task and about the naturalness of the stimuli. Globally, the meaning of the labels chosen for defining the semantic scales was clear to them, and the stimuli sounded natural. The experiment lasted 20 minutes.

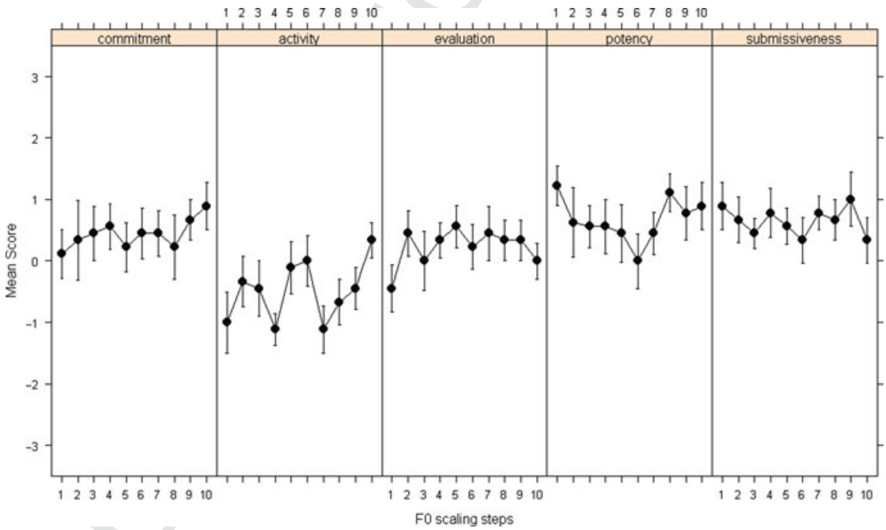
3.1.4 Participants

Nine listeners participated in the experiment, three females and six males between 30 and 50 years of age without known hearing disorders. The listeners were all brought up in Naples and spoke standard Italian with a Neapolitan accent. None of them was a specialist in linguistics.

3.2 Results

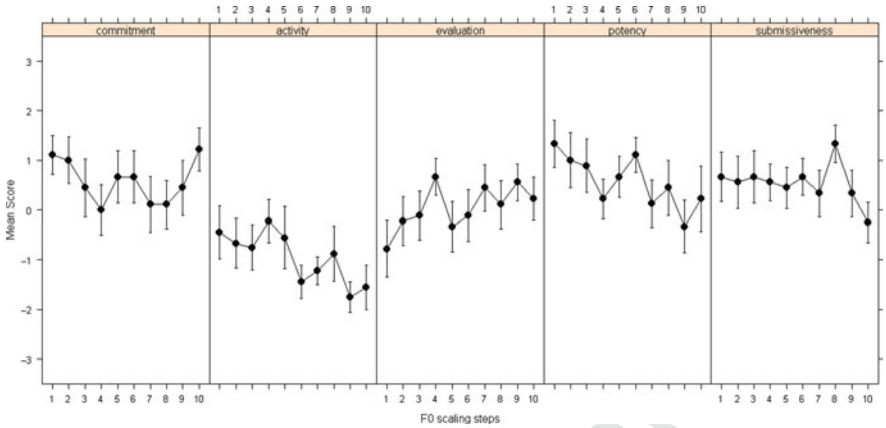
Figures 7 and 8 show results for stimuli in the PREN and in the AP condition, respectively. In both graphs, mean judgment scores are plotted across the 10 scaling steps separately for the five semantic scales. The statistical relevance of such results was also tested by a series of linear mixed models, separately run for the two gates conditions and for each semantic scale. Specifically, the mean judgment scores were the dependent variable, while the AP scaling manipulation was the fixed factor and listeners constituted the random one ($pMCMC < .01$).

Let us focus first on results for the PREN condition. Remember that in such condition, stimuli only contained the prenuclear rise (LH*) and some portion of the following f_0 movement towards the AP tone. As we can notice from Fig. 7, the judgment scores did not vary with the scaling manipulation: they were centralized around the '0' level, meaning that listeners were not able to interpret these stimuli neither as 'questioning' or 'asserting'. This was expected, since in this experiment possible effects of phonetic factors (such as speech rate) were neutralized by employing only one base sentence (a statement). The regression analyses confirmed that the manipulation of the AP height did not significantly affect listeners' responses for any of the five scales ($pMCMC > .01$). This also means that the differences in the steepness of the short f_0 movements following the prenuclear peak were not able to affect subjects' responses.



This figure will be printed in b/w

Fig. 7 Mean judgments scores for the five semantic scales ('commitment', 'activity', 'evaluation', 'potency' and 'submissiveness') in the PREN (upper panel) and AP (lower panel) conditions. Results are pooled for all listeners



This figure will be printed in b/w

Fig. 8 Mean judgments scores for the five semantic scales ('commitment', 'activity', 'evaluation', 'potency' and 'submissiveness') in the PREN (upper panel) and AP (lower panel) conditions. Results are pooled for all listeners

Fig. 8, where scores for stimuli in the AP condition are presented, offers a partially different picture. Remember that these stimuli contained scaling information relative to the AP tone. Specifically, the mean judgments score progressively decreases as the height of the AP tone increases in two out of five scales, the 'activity' and the 'potency' one. In the 'activity' scale, the mean score at the left extreme of the continuum (H1) was close to zero (−0.4), thus indicating that when the AP scaling is very low, the speaker sounds 'neutral'. Mean score decreases in the following scaling steps, so that it reaches −1.6 at the right extreme of the continuum (H9 – H10). This suggests that raising the f_0 height of the AP tone shifts judgments from 'neutral' to 'slightly involved'. In the 'potency' scale, the mean judgments values decrease from 1.3 (H1) to 0.2 (H9) and −0.3 (H10). This suggests that when the scaling of the AP tone is raised, there is a shift from the 'slightly certain' to the 'neutral' interpretation. Such a linear decrease in mean scores is significant for both the 'activity' [t = −3.1, $pMCMC < .01$] and the 'potency' [t = −2.64, $pMCMC < .01$] scales.

An additional observation needs to be made. One might wonder why responses for the 'activity' and 'potency' scales do not drastically vary as a function of the scaling manipulation. This might be due to many factors. For example, we might think that the difference in the phonological specification of the AP tone is a secondary correlate of the contrast between questions and statements, whose 'informational weight' is less important than the one carried by nuclear accent alignment. Another explanation (which does not exclude the first one) is that the judgments score has been affected by the nature of the

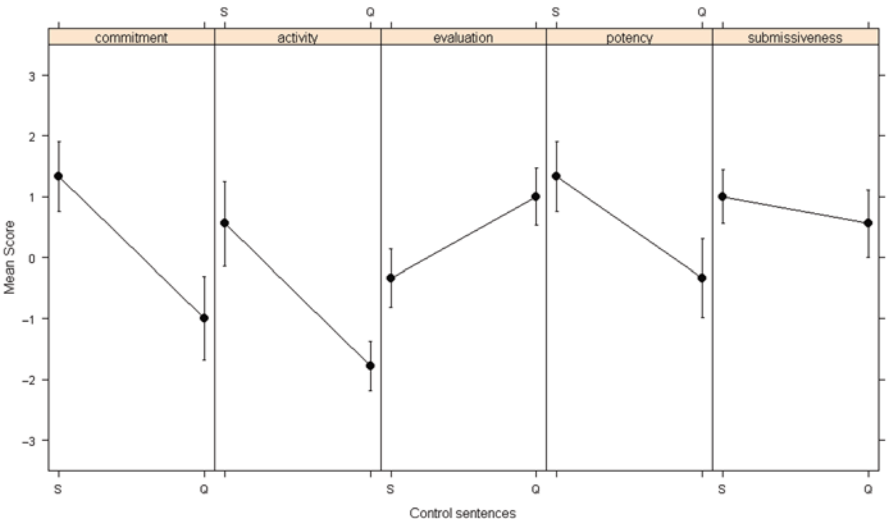


Fig. 9 Mean judgments scores for the five semantic scales ('commitment', 'activity', 'evaluation', 'potency' and 'submissiveness') in the NUCL condition stimuli set. Results are pooled for all the listeners

stimuli. In fact, the use of gated stimuli could have caused an increase in the degree of uncertainty in subjects' judgments. However, what is important here is that, despite its small magnitude, the effect of AP scaling is significant in both a linguistic and a paralinguistic scale. This suggests that the contrast between L_{AP} and H_{AP} is systematically employed by Neapolitan listeners even when the nuclear accent information is not available.

Fig. 9 shows results for control sentences, containing both the prenuclear and the nuclear contours. Mean judgments scores (y axis) are plotted against the two stimuli chosen as representative of the contrast between statements (S, left) and questions (Q, right). Such stimuli differed both in the specification of the AP tone (L_{AP} vs. H_{AP}) and in that of the nuclear accent (L + H* vs. L* + H). The difference in mean scores between these two stimuli is larger than the one found between the corresponding ones in the AP condition, probably reflecting the higher confidence of listeners in rating stimuli as complete utterances containing the nuclear pitch accent. Specifically, in the 'commitment' scale, the mean score was 1.3 for the statement stimulus (the speaker sounds 'quite incontestable') and −1 for the question stimulus (i.e., the speaker sounded 'slightly contestable'). Thus, in this scale there was a two-step difference between the two stimuli. In the 'activity' scale, such a score reached 0.5 in the statement stimulus and −1.7 in the question. There is thus a one-step difference for this scale: the speaker is judged 'neutral' when the AP tone is

L_{AP} and the nuclear accent is L + H*, but he is 'slightly involved' when the AP tone is H_{AP} and the nuclear accent is L* + H. Mixed models confirmed that the difference between the two stimuli is significant both in the 'activity' [$t = -2.99$, $pMCMC < .01$] and in the 'commitment' scale [$t = -2.33$, $pMCMC < .01$]. Though our results are based on relatively few observations (i.e., 18 observations for each scale in the NUCL condition), it is noteworthy that they are very similar to those found for the AP condition. Specifically, stimuli in the AP and NUCL conditions affect subjects' responses both in the linguistic ('potency' and 'commitment') and paralinguistic ('activity') scales. Note also that the direction of the effect is similar between the two conditions, thus suggesting that the meaning carried by the AP tone is congruent with that carried by the nuclear accent. This last point is further discussed in the next section.

3.3 Discussion

In this experiment, we tested whether there is a scaling effect due to the presence of the L_{AP} and the H_{AP} tones, which could be exploited by Neapolitan listeners to distinguish statements from questions. Preliminary results from a semantic differential task revealed that the AP tone carries both a linguistic and paralinguistic meaning. However, we also found that the modification of the semantic scales differs depending on the specification of the AP tone. First, in stimuli containing only the prenuclear rise and the AP tone, low AP values convey the information that speaker is more certain about the content of his message. The degree of 'certainty' progressively decreases as the AP tone height increases. Note that the idea of 'certainty' is strictly linked to that of 'assertion' according to the linguistic interpretation of the Frequency Code. This result is also congruent with the one found for stimuli also containing the nuclear L + H* accent, ie, the typical nuclear accent for Neapolitan narrow focus statements. Specifically, in stimuli containing also the L + H* accent listeners' judgments were shifted towards 'incontestability', meaning that the speaker sounds less prone to revise the propositional content of his message. The idea of 'incontestability', which might be related to that of 'assertion', is also expressed by terminal low tones/falling contours in other languages (Gussenhoven 2002). The fact that, in Neapolitan, such an idea is conveyed by stimuli containing the L + H* accent is not surprising. We know that the contrast between questions and statements is often associated to a contrast between high and low pitch, whose instantiation can be different across languages (Gussenhoven 2002, 2004). In particular, in Neapolitan, low pitch in statements might be signaled by the earlier alignment of the nuclear peak (L + H*), while high pitch in questions might be signaled by the nuclear peak delay (L* + H). As a consequence, the fact that Neapolitan speakers pronouncing a nuclear L + H* sound more incontestable is in line with the Gussenhoven's hypothesis concerning the grammaticalisation of the Frequency Code.

The interpretation of the results for the H_{AP} tone in the AP condition is more problematic, since this tone does not seem to modify responses' score on the linguistic scales. One might wonder whether this result points to the absence of an AP tone altogether in the question modality. In other words, the 'salient' information in the prenuclear region would consist merely in the presence vs. absence of an L_{AP} tone: when such a tone is present, Neapolitans would hear more statements, while its absence would induce listeners to perceive more questions, or at least 'non-assertive' utterances.

However, this hypothesis is dispreferred for two reasons. First, previous acoustic results (Petrone and D'Imperio 2008; D'Imperio and Petrone 2008; Petrone 2008) revealed that, though the region following the prenuclear accent is characterized by a difference in the shape and slope between the two intonation modalities, the inflection point of the curve is realized at similar temporal location (around the end of the first prosodic word). This suggests that a comparable tonal event is present in both questions and statements, that is, the inflection point of the curve might be due to the insertion of a tone signalling the end of the accentual domain in both modalities.

Moreover, in Experiment II we found that, while there was no modification of listeners' judgments in the PREN condition, the presence of the H_{AP} tone in the AP condition appears to be exploited for getting information about the speaker's emotional state, in that the speaker sounds more involved in the discourse. This attitudinal meaning is also carried by stimuli containing the nuclear $L^* + H$, i.e., the typical nuclear accent for Neapolitan yes/no questions. Namely, the idea of 'emotional involvement' in the discourse is stronger for stimuli containing both the H_{AP} tone and the $L^* + H$ accent than in stimuli only containing the H_{AP} tone. This stronger effect can be due to the fact that listeners are more confident in rating stimuli containing a nuclear pitch accent than those without it. However, the nuclear accent also carries linguistic information, which is related to the question modality. In fact, the stimulus chosen as representative of Neapolitan questions conveyed the idea that the speaker is 'more contestable'.

But why is there a discrepancy in the results obtained for L_{AP} (conveying a linguistic information) and H_{AP} (conveying a paralinguistic one)? The question is still open, and there are, in our view, different possible explanations. First, the H_{AP} tone might be an affective morpheme. However, though an affective morpheme has been attested in at least one language, Dutch (Grabe et al. 1998; but see Gussenhoven 2004 for a different explanation), this hypothesis seems implausible for Neapolitan Italian since it would imply a functional difference for two tones in the same structural position.

Another possible hypothesis is that the presence of the L_{AP} is a stronger cue for statements than the H_{AP} for questions. This hypothesis stems from two considerations. First, it has been noted that, cross-linguistically, statement intonation is less commonly characterized by non-falling intonation than question modality by non-rising intonation (Gussenhoven 2002). Therefore, it is possible that falling intonation tends to be associated to statements more often than rising intonation to questions. This effect can be even stronger in the case of Neapolitan, where yes/no questions are mainly cued by the late alignment of the nuclear accent. Such a hypothesis is strengthened by results from a pilot study in German (Petrone and Niebuhr 2009). We already know that in German, the distinction between yes/no questions and statements is mainly indicated by the terminal fall (L%) vs. rising intonation (H%). However, similarly to Neapolitan, we found that the f_0 fall after the prenuclear rise is shallower and it has a convex shape in questions, while it is steeper in statements. Such a difference in the prenuclear region seems to be already sufficient for the perception of the intonation contrast also in that language: listeners' judgments were significantly shifted towards the 'question' interpretation for stimuli containing up to the convex, shallow fall; and they were shifted towards the 'statement' interpretation for stimuli containing up to the steep fall. Also, similar to Neapolitan, question scores further increased when the nuclear pattern was available. This suggests that the nuclear configuration is an important cue for German questions, though, as in Neapolitan, it is not necessary ('question' responses being well above the chance level even in absence of the nuclear pattern).

Note also that cross-linguistic studies have shown language-specific differences in the perceptual association of high pitch with questions (see Gussenhoven 2002 for a short summary). So, for example, Gussenhoven and Chen (2000) found that, though Dutch, Chinese and Hungarian listeners tended to associate higher peaks, higher end pitch and later peaks with question modality, in Hungarian higher peaks attracted more 'question' judgments. In fact, in this language, differently by the other two, high peak is used to signal questions. Similarly, we might think that German speakers are more prone to rely on high pitch for interrogativity than Neapolitans, since high pitch is also used as main cue for questions. This will be tested in future research.

Another possible explanation is that the way in which we created stimuli for Experiment II was not appropriate to capture the difference in the prenuclear contour between the two intonation modalities. In line with the AM theory, such a difference was interpreted as due to the insertion of two static tones, L_{AP} and H_{AP}. However, previous acoustic studies (Petrone and D'Imperio 2008; D'Imperio and Petrone 2008) suggest that dynamic factors might play a role in defining such a contrast. Specifically, we found that in questions, the f_0 region following the prenuclear peak has a shallow slope, thus assuming a convex shape. Such characteristics have not been taken into account for our stimuli, which were created by simply raising the melodic value of a single point in the contour (the L_{AP} tone) and by connecting it to the neighbouring points through linear interpolations. Therefore, our stimuli might not entirely reflect some crucial dynamic difference between the two intonation modalities. If this is true, results of Experiment I might be also reinterpreted.

Remember that in Experiment I, no differences were found between the PREN and the AP conditions for question base stimuli. In fact, stimuli in the PREN condition already contained some portion of the convex shape typical for questions. If Neapolitan listeners capitalize on the dynamic properties of the fall (and not on the availability of a single point, the H_{AP} tone), they could have been exploiting such cues already in the PREN condition. The perceptual impact of such dynamic properties will be better explored in future investigation.

4 Conclusion

In Neapolitan Italian, the distinction between yes/no questions and narrow focus statements appears to be signaled both by a contrast between nuclear accent types (L* + H vs. L + H*) and by the presence of an AP tone (H_{AP} vs. L_{AP}) realized in the f0 prenuclear region. While previous studies have reported that the alignment of the nuclear accent is a robust perceptual cue for intonation modality, results from our identification and semantic differential tasks suggest that Neapolitans are able to differentiate questions and statements well before the nuclear accent is perceived. First, differences in the perception of the two modalities seem to be at work already in very early portions of the utterance, i.e., even before AP target location. Moreover, differences in AP target scaling significantly affected listeners' judgments. This also calls for a better understanding of the semantic weight carried by the prenuclear contour, as well as the interaction between the independent contribution of the prenuclear and nuclear contours in conveying tune meaning.

Acknowledgments This article developed material presented at the Conference *TIE3* and we are very grateful to the audience of the conference. Thanks also to Sue Hertz and Lisa Selkirk for fruitful discussions and to reviewers for comments to an earlier version of the paper. Thanks also to Dr. Cinzia Citraro for technical help. All errors are of course ours.

References

André, Carine, Alain Ghio, Christian Cavé, and Bernard Teston. 2007. *PERCEVAL: PERCeption EVALuation Auditive and Visuelle* (version 3.0.4). http://www.lpl.univaix.fr/~lpl/dev/perceval

Baayen, R. Harald. 2008. *Analyzing Linguistic Data: A Practical Introduction to Statistic*. Cambridge: CUP.

Bartels, Christine. 1999. *Towards a Compositional Interpretation of English Statement and Question Intonation*. New York: Garland.

Brinckmann, Caren, and Ralf Benzmueller. 1999. The relationship between utterance type and F0 contour in German. In *Proceedings of Eurospeech 1999*, Budapest, vol. 1, 21–24.

Chen, Aoju. 2003. Reaction Time as an indicator of discrete intonational contrast in English. In *Proceedings of Eurospeech 2003*, Geneva, 97–100.

D'Imperio, Mariapaola. 2000. *The Role of Perception in Defining Tonal Targets and their Alignment*. Ph.D. Thesis, The Ohio State University.

D'Imperio, Mariapaola, Francesco Cangemi, and Lisa Brunetti. 2008. The phonetics and phonology of contrastive topic constructions in Italian. Poster presented at the *3rd Conference on Tone and Intonation in Europe*, September 15–17, Lisbon, Portugal.

D'Imperio, Mariapaola, and David House. 1997. Perception of Questions and Statements in Neapolitan Italian. In *Proceedings of Eurospeech '97*, Rhodes, Greece, vol. 1, 251–254.

D'Imperio, Mariapaola, and Caterina Petrone. 2008. Is the Clitic Group tonally marked in Italian questions and statements? Poster presented at the *11th International Conference on Laboratory Phonology*, June 30–July 2, Wellington, New Zealand.

Dombrowski, Ernst. 2003. Semantic features of accent contours: effects of F0 peak position and F0 time shape. In *Proceedings of the International Conference of Phonetic Sciences,* 1217–1220, Barcelona, Spain.

Grabe, Esther, Carlos Gussenhoven, Judith Haan, Erwin Marsi, and Brechtje Post. 1998. Preaccentual pitch and speaker attitude in Dutch. *Language and Speech* 41: 63–85.

Grice, Martine. 1995. Leading tones and downstep in English. *Phonology* 12: 183–233.

Grice, Martine, Mariapaola D' Imperio, Michelina Savino, and Cinzia Avesani. 2005. Towards a strategy for labelling varieties of Italian. In Sun-Ah Jun, (ed.) *Prosodic Typology and Transcription: A Unified Approach*, 53–83. Oxford: Oxford University Press.

Grosjean, François. 1980. Spoken word recognition processes and the gating paradigm. *Perception and Psychophysics* 28: 267–283.

Gussenhoven, Carlos. 1984. *On the Grammar and Semantics of Sentence Accents*. Dordrecht: Foris Publications.

Gussenhoven, Carlos. 2002. Intonation and interpretation: Phonetics and phonology. In Bernard Bel and Isabelle Marilier, (eds.) *Proceedings of Speech Prosody,* 45–57, Aix-en-Provence.

Gussenhoven, Carlos. 2004. *The Phonology of Tone and Intonation*. Cambridge: Cambridge University Press.

Gussenhoven, Carlos, and Aoju Chen. 2000. Universal and language-specific effects in the perception of question intonation. International Conference on Spoken Language Processing 6(II): 91–94.

Haan, Judith. 2001. S*peaking of Questions. An Exploration of Dutch Question Intonation. LOT Dissertation* Series 52, LOT, Utrecht.

Ladd, D. Robert. 2008. *Intonational Phonology* (II edition). Cambridge: Cambridge University Press.

Lahiri, Aditi, and William Marslen-Wilson. 1991. The mental representation of lexical form: A phonological approach to the recognition lexicon. *Cognition* 38: 254–294.

Ohala, John. 1983. Cross-language use of pitch: an ethological view. *Phonetica* 40: 1–18.

Marandin, Jean-Marie. 2006. Contours as Constructions. In Doris Schoenefeld (ed.) *Constructions All Over; Case Studies and Theoretical Implications*, Ms available at http://www.constructions-online.de/articles/specvol1

Marandin, Jean-Marie, Claire Beyssade, Elisabeth Delais-Roussarie, Jenny Doutjes, Anne Rialland, and Michel de Fornel. 2004. The meaning of final contours in French. Ms available at http://www.llf.cnrs/ Gens/ Marandin

Moulines, Eric, and Francis Charpentier. 1990. Pitch-synchronous waveform processing techniques for text-to-speech synthesis using diphones. *Speech Communication* 9: 453–467.

Niebuhr, Oliver. 2007. The signalling of German rising-falling intonation categories-the interplay of synchronization, shape, and height. *Phonetica* 64: 174–193.

Niebuhr, Oliver, and Klaus Kohler. 2004. Perception and cognitive processing of tonal alignment in German. In *Proceedings of the TAL Conference*, 155–158.

Petrone, Caterina. 2008. *Le rôle de la variabilité phonétique dans la représentation des contours intonatifs et de leur sens*. Ph.D. Thesis, Université de Provence, France.

Petrone, Caterina. 2010. At the interface between phonetics and pragmatics: Non-local F0 effects on the perception of Cosenza Italian tunes. In *Proceedings of Speech Prosody 2010*.

Petrone, Caterina, and Mariapaola D'Imperio. 2008. Tonal structure and constituency in Neapolitan Italian: Evidence for the accentual phrase in statements and questions. In Plinio A. Barbosa, Sandra Madureira, and César Reis, (eds.) *Proceedings of the 4th Conference on Speech Prosody,* 301–304, Campinas, Brazil.

Petrone, Caterina, and Oliver Niebuhr. 2009. The role of the prenuclear f_0 region in the perception of German questions and statements. Talk presented at *9th Conference on Phonetics and Phonology in Iberia*. Presentation available at http://www.linguistik.uni-kiel.de/Niebuhr_index.html

Petrone, Caterina, and D. Robert Ladd. 2007. Sentence domain effects on tonal alignment in Italian. In Jürgen. Trouvain, and William. J. Barry, (eds.) *Proceedings of the 16th International Conference of Phonetic Sciences,* vol. 2, 1253–1256, Saarbruecken, Germany.

Pierrehumbert, Janet B. 1980. *The Phonology and Phonetics of English intonation*. Ph.D. Thesis, MIT.

Pierrehumbert, Janet B., and Mary E. Beckman. 1988. *Japanese Tone Structure*. Cambridge, MA: MIT Press.

Pierrehumbert, Janet B., and Julia Hirschberg. 1990. The meaning of intonational contours in the interpretation of discourse. In Philip R. Cohen, Jerry Morgan, and Martha E. Pollack, (eds.) *Intentions in Communication*, pp. 271–311. Cambridge, Massachusetts: MIT Press.

Pierrehumbert, Janet B., and Shirley Steele. 1987. How many rise-fall-rise contours?. *Proceedings of the 11th International Conference of Phonetic Sciences (IChPS 1987)*, 145–148.

Pinheiro, José C., and Douglas Bates. 2000. *Mixed-Effects Models in S and S-Plus*. *Statistics and Computing Series*. New York, NY: Springer-Verlag.

Prieto, Pilar, Mariapaola D'Imperio, and Barbara Gili-Fivela. 2005. Pitch accent alignment in Romance: primary and secondary associations with metrical structure. *Language and Speech* 48: 359–396.

R Development Core Team. 2008. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org

Rathcke, Tamara, and Jonathan Harrington. 2010. The variability of early accent peaks in Standard German. In Cécile Fougeron, Barbara Ka Hnert and Mariapaola D'Imperio, (eds.) *Papers in Laboratory Phonology 10*, Berlin: Mouton de Gruyter.

Savino, Michelina, and Martine Grice. 2007. The role of pitch range in realising pragmatic contrasts – The case of two question types in Italian. In Jürgen. Trouvain, and William. J. Barry, (eds.) *Proceedings of the 16th International Conference of Phonetic Sciences (ICPhS 2007)*, 1037–1040, Saarbruecken, Germany.

Savino, Michelina, and Martine Grice. 2008. Reaction time in the perception of intonational contrast in italian, paper presented at *TIE3 Conference*, Lisbon 15–17 Sept. 2008.

Schneider, Katrin, Britta Lintfert, Grzegorz Dogil, and Bernd Mobius. 2006. Phonetic grounding of prosodic categories. In Stefan Sudhoff, Denisa Lenertová, Roland Meyer, Sandra Pappert, Petra Augurzky, Ina Mleinek, Nicole Richter, and Johannes Schließer (eds.), *Methods in Empirical Prosody Research*, 335–361. Berlin, Germany: De Gruyter.

Steedman, Mark. 2003. Information-structural semantics for English intonation. (Paper to LSA Summer Institute Workshop on Topic and Focus, Santa Barbara July 2001). Ms available at www.cogsci.ed.ac.uk/~steedman/.

't Hart, Johan. 1981. Differential sensitivity to pitch distance. *The Journal of the Acoustical Society of America* 69: 811–821.

Uldall, Elizabeth T. 1960. Attitudinal meanings conveyed by intonation contours. *Language and Speech* 3: 223–234.

van Heuven Vincent J., and Judith Haan. 2000. Phonetic correlates of statement versus question intonation in Dutch. In Antonis Botinis, (ed.) *Intonation. Analysis, Modelling and Technology*, 119–143. Dordrecht: Kluwer.

van Heuven, Vincent J., and Ellen van Zanten. 2005. Speech rate as a secondary prosodic characteristic of polarity questions in three languages. *Speech Communication* 47: 87–99.

The Role of Pitch Cue in the Perception of the Estonian Long Quantity

Pärtel Lippus, Karl Pajusalu, and Jüri Allik

1 Introduction

A central feature of Estonian word prosody is the three-way quantity distinction. The domain of quantity is a primary stressed disyllabic foot. The distinction of short (Q1), long (Q2) and overlong (Q3) quantity degrees can be realized by vowels, syllable-medial consonants or both as well as combinations of diphthongs and consonant clusters (see Table 1).

Phonologically, this distinction occurs only in the stressed syllable; there is no phonological length opposition in the following unstressed syllable (Viitso 2003). Phonetically, the quantity distinction is realised by the durational ratio of the segments in the trochaic foot; that is, the quantity degrees are not perceivable if the second syllable is not present (Eek and Meister 2003, 2004). Due to a certain amount of foot isochrony (Lehiste 2003, Nolan and Asu 2009), the second syllable duration compensates for variation in the first syllable duration. In Q1, a short open stressed syllable is followed by a long (half-long) syllable. In Q2, a long stressed syllable is followed by a short syllable. Finally, in Q3 an extra-long stressed syllable is followed by an extra-short syllable. Ilse Lehiste (1960, 1997, 2003) described the quantity degrees as syllable duration ratios: 2/3 in Q1, 3/2 in Q2 and 2/1 in Q3 (similar syllable duration ratios are presented also by Liiv 1961 and others). It has been argued, that the duration ratio of syllables is not possible feature for the perception of quantity. Rather, the duration ratios of neighbouring sound segments is a better way to describe the quantities (Traunmüller and Krull 2003; Eek and Meister 2003, 2004) However, all agree that the duration of segments in syllable rhymes is more important for the quantity distinction and that the duration of syllable onsets carry most of the information about speech rate.

One account of Estonian intonational phonology can be found in Asu (2004). The most common pitch accent in Estonian is H* + L in which case

P. Lippus (✉)
Institute of Estonian and General Linguistics, University of Tartu, Ülikooli 18, 50090 Tartu, Estonia
e-mail: partel.lippus@ut.ee

S. Frota et al. (eds.), *Prosodic Categories: Production, Perception and Comprehension*, Studies in Natural Language and Linguistic Theory 82,
DOI 10.1007/978-94-007-0137-3_10, © Springer Science+Business Media B.V. 2011

Table 1 Possible combinations of sound segments that demonstrate the quantity opposition

		Q1	Q2	Q3
Vocalic	Single sound	[sɑ.tɑˑ] 'hundred'	[sɑː.tɑ] 'send!' sg2 imp	[sɑːː.tɑ̆] 'to get' inf
	Diphthong	*	[vɤi.tɑ] 'win!' sg2 imp	[vɤiː.tɑ̆] 'to oil' inf
Consonant	Single sound	[kɑ.tɑ.] 'slingshot'	[kɑt.tɑ] 'cover!' sg2 imp	[kɑtː.tɑ̆] 'to cover' inf
	Consonant cluster	*	[kɑs.tɑ] 'water!' sg2 imp	[kɑsː.tɑ̆] 'to water' inf
Both	Single sounds	*	[sɑːt.te] 'get' pl2	[sɑːtː.tĕ] 'broadcast' sg gen
	Diphthong and cc	*	[mɑit.se] 'taste' sg nom	[mɑitː.sĕ] 'taste' sg gen

the pitch contour of accented words is normally characteristic of the three quantities. As a secondary feature pitch plays an important role in signalling the three-way quantity distinction of Estonian. In the previous studies, the peak, the pitch range, the duration of the rising or high part, the duration of the falling part following the peak, and the over-all characteristics of the pitch pattern have been discussed. According to the traditional view, the pitch contour is rising-falling in all quantity degrees, but in Q1 and Q2 words the position of the peak is at the end of the first syllable, while in Q3 words it is in the beginning of the first syllable (Lehiste 1960; Liiv 1961; Remmel 1975).

By looking only at the first syllable, the pitch contour has been described as 'level' or 'rising' in Q1 and Q2 but 'falling' in Q3 words (Liiv 1961). For Q1 and Q2 the term 'rising-returning' has also been used, meaning that the pitch rises in beginning but returns to the initial level at the end of the syllable. In the second syllable the pitch falls abruptly in Q1 and Q2, but more gently in Q3 (Lehiste 1960).

If the overall pitch contour of the word is considered, it can be described as falling or rising-falling in all quantity degrees, but the contour can also be viewed as having a high part, a falling part, and a low part (Lehiste 2003; Asu et al. 2009). The latest corpus-based study of spontaneous speech by Asu et al. (2009) finds that the most typical pitch contour of disyllabic feet is falling rather than rising-falling. In all quantity degrees the word begins with a high plateau where the pitch is considerably level until the point where the pitch begins to fall (Asu et al. 2009).

The main difference in the overall pitch contour between the quantity degrees is the distribution of the high pitch and fall, or the location of the peak. In Q1 and Q2 words the peak is located at 3/4 of the first syllable duration, but in Q3 words the peak is at 1/4 of the first syllable duration (Liiv 1961; Remmel 1975). In Q1 and Q2 words there is a high pitch plateau before a fall at the end of the first syllable, but in Q3 words the high pitch plateau turns to a fall in the first half of the first syllable. However, the absolute duration from the beginning of the syllable to the turning point is not significantly different between quantity degrees (Asu et al. 2009).

No significant difference in pitch range between the quantity degrees has been found (Lehiste 1960; Liiv 1961). Liiv describes the pitch range as follows: in Q1 the rise is a minor third (3 semitones); in Q2 the rise is a major third (4 st) or a fourth (5 st) followed by a major second (2 st) of fall; in Q3 the pitch rises a major second (2 st) and falls by a fourth (5 st) or a major third (4 st; Liiv 1961). The over-all pitch range of the foot does not differ between the quantity degrees. However, due to the different location of the turning point in the first syllable, the pitch ranges less than 0.5 st in Q1 and Q2, but about 2 st in Q3 (Asu et al. 2009).

A study of the interaction of rising (L* + H) intonation with the tonal characteristics of the quantities (Asu and Nolan 1999) showed that in the case of rising intonation the rise in all quantities starts just before the end of the first vowel and not earlier in the case of Q3. In the case of low accentuation patterns, where there is a fall from a high unaccented syllable to a low accented syllable (H + L*), the F0 contour is realized as flat and low throughout the pitch accent or is falling right at the beginning of the first syllable independent of the quantity (Asu and Nolan 2007). However, such pitch patterns seem to be rather rare in Estonian; among 348 analysed words, Asu et al. (2009) found only a couple of such tokens in their corpus.

Perception studies have shown that the pitch cue is crucial for distinguishing Q2 and Q3. Lehiste (1970) used synthesized stimuli with various S1/S2 duration ratios and three different pitch contours: level pattern (monotone 120 Hz), step-down pattern (S1 120 Hz, S2 80 Hz; typical for Q2), and falling pattern (S1 120-80, S2 monotone 80 Hz; typical for Q3). The results showed that with flat F0 the judgement of the stimuli as Q2 or Q3 depended mostly on the temporal structure. If the pitch contour was falling, Q3 was favoured, but if the pitch was the step-down contour, Q3 was not recognized even if the temporal structure was typical for a Q3 word. Q1 was discriminated from Q2 with all the pitch patterns.

A similar experiment was conducted by Eek (1980a and 1980b) using re-synthesized natural speech stimuli wherein the duration of V1 and/or V2 was manipulated. The F0 of the base words was rising in Q1 and Q2 words and falling in Q3, and it remained unmodified. The results showed that if only the temporal structure was modified, Q1 and Q2 could be converted into each other. It was not possible to obtain an acceptable Q3 from a Q1 or a Q2 word by modifying the duration, nor could a natural Q1 or a Q2 word be obtained from a Q3 word (Eek 1980a). Q3 was perceived from Q1 and Q2 words only if also the pitch was modified from rising to falling and the S1/S2 duration ratio was typical for Q3 (Eek 1980b).

Lippus et al. (2007, 2009, 2009) studied the perception of Estonian quantity degrees by native and non-native speakers. Natural speech words were re-synthesized, wherein only the duration of V1 was modified. The F0 contour was slightly rising in V1 and falling in V2 in Q1 and Q2 words, and generally falling in Q3. Among the native subjects, two distinct groups were identified on the basis of their dialectal background and the significance of the pitch cue to their perception of the quantity. The subjects from central and western Estonian dialect regions did not perceive Q3 in tokens derived from Q1 or in Q2 words but when the token was re-synthesized from a Q3 word, the subjects perceived all the quantity degrees on the basis of the durational structure. The subjects from eastern and southern Estonian dialect areas perceived all the quantity degrees on the basis of the durational structure of the stimuli and were not influenced by the pitch contour (Lippus and Pajusalu 2009).

Based on the results from Lehiste (1970), Lehiste and Danforth (1977) present a hierarchy of phonetic cues for the perception of Estonian quantities where pitch cue follows after the duration of V1 (or S1 rhyme in case of consonant quantity). Within the whole foot, the tonal peak is at the end of the nucleus of the stressed syllable in Q1 and Q2, but falls noticeably in the unstressed syllable. In Q3, the pitch starts falling in the first half of the stressed syllable and the fall continues in the unstressed syllable. Based on all of her perception tests, Lehiste concludes that the quantity opposition is binary; syllable ratios discriminate short from long, but for the discrimination of long and overlong, the pitch is vital (Lehiste 1997, 2003). Eek concludes that Estonian quantity is durational-accentual: Q1 vs. Q2 is an opposition of short and long, but Q2 vs. Q3 is an opposition of lax-long and tense-long (Eek 1980b).

The Estonian quantity system has developed from a short-long opposition to a three-way opposition as a result of a number of language changes, including apocope and syncope (Kask 1972). Lehiste stated that the over-length of Q3 arose due to compensatory lengthening; the overlong syllable of Q3 now carries the pitch contour of what was formerly a disyllabic sequence (Lehiste 1978; Lehiste 2003). In another experiment Lehiste demonstrated that with an earlier peak in a rising-falling pitch contour a signal is perceived to be longer than a signal of same length but with a later peak (Lehiste 1976).

In all previous experiments where the pitch has been manipulated, it has been done in combination with the manipulation of the temporal structure. In this paper we test the role of the tonal component in distinguishing Q2 and Q3 words by changing the pitch of a Q2 word without changing the duration. We use various synthesized pitch contours for establishing the most typical F0 for Q3.

2 Experiment 1

2.1 Materials and Methods

The stimuli were created using the PRAAT-software (Boersma and Weenink 2007). The Q2-word *saada* [sɑːtɑ] ‘send!’ was recorded when pronounced in isolation by a male speaker. The syllable duration ratio for the word was 1.4 (according to Lehiste (1997) the typical V1/V2 ratio for Q2 is 1.5 and for Q3

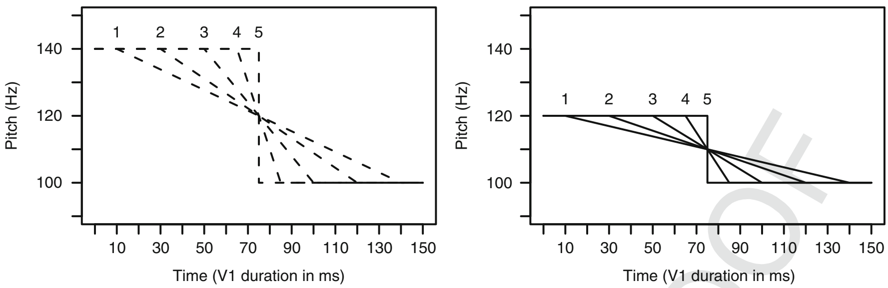


Fig. 1 Schematic pitch contours of the V1 of the stimuli. The starting point of the fall is marked with the stimulus number. *Left*: Set 1 with the pitch range from 100 Hz to 140 Hz. *Right*: Set 2 with the pitch range from 100 Hz to 120 Hz. In V2 the pitch continued at 100 Hz in all stimuli

more than 2). Stimuli were re-synthesized by changing the F0 contour of the word while leaving the duration un-manipulated.

Two sets of five stimuli were created, where the duration of the pitch fall was altered. The locus of the fall was always in the middle of V1. The duration of the fall was varied in five steps from 130 ms in the first stimulus to 0 ms in the last stimulus, see Fig. 1. In the first set, the F0 ranged from 100 Hz to 140 Hz (about 6 semitones; Set 1), and in the second set the F0 varied from 100 Hz to 120 Hz (about 3 semitones; Set 2).

The results of 22 native Estonian speakers (8 male, 14 female; age 20–61 years, average being 33) are reported. The test subjects were students and faculty members of the University of Tartu and the Tallinn University of Technology. A forced-choice perception experiment was carried out using Praat. The stimuli were presented to the test subjects in two blocks of 5 stimuli with 10 repetitions in random order (i.e. $2 \times 5 \times 10 = 100$ stimuli in total). The subjects were told that they will hear synthesized Q2 and Q3 words and were instructed to think about the meaning of the words: 'send!' in case of Q2 and 'to get' in case of Q3. The subjects listened to the stimuli using headphones and had to decide whether they heard a Q2 word or a Q3 word, and click a button on the computer screen, labelled [2] and [3] accordingly.

2.2 Results

The results of the test divide the subjects into two groups. The main group (Group 1) was formed by 16 subjects who judged the stimuli according to the pitch pattern and gave both Q2 and Q3 responses in both stimuli sets. The 6 subjects in the second group (Group 2) judged all the stimuli in Set 1 as Q3 and in Set 2 as Q2. No differences in social or regional background between the two groups were found. The variation appears to be in different perceptional abilities. The results of both groups are presented in Table 2 and Fig. 2.

Table 2 The percentage of Q3 responses in Experiment 1

	Stimulus 1	Stimulus 2	Stimulus 3	Stimulus 4	Stimulus 5
Set 1 Group 1	41%	64%	85%	84%	77%
Set 1 Group 2	83%	97%	98%	98%	98%
Set 2 Group 1	27%	41%	54%	58%	51%
Set 2 Group 2	2%	8%	5%	18%	7%

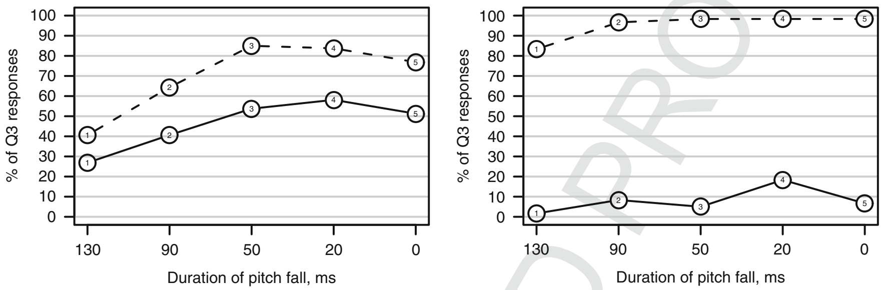


Fig. 2 Percentage of Q3 responses in Experiment 1. Set 1 is connected with *dotted lines*, Set 2 with *solid lines*. *Left*: the results of Group 1. *Right*: Results of Group 2

For Set 1 both groups of subjects more frequently selected Q3. Group 1 had the highest Q3 judgment rate when presented with Stimuli 3 and 4. In Stimulus 3, the duration of the high plateau, the fall, and the low plateau were all 1/3 of the V1 duration. In Stimulus 4, the duration of the pitch fall was 20 ms (13% of the V1 duration) and the durations of the high and low plateaus were 65 ms (43% of the V1 duration). Stimuli 1 and 2 which had the longer pitch fall and the shorter plateaus, received the lowest Q3 responses (41% and 64% respectively). Stimulus 5 which had the rapid pitch fall at 50% of the V1 duration was most often perceived as Q3, but the rate of Q3 responses was again lower than that for the Stimuli 3 and 4. Group 2 of subjects only selected Q3 for all Set 1 stimuli.

In Set 2 with a pitch range of 3 st, Stimulus 4 was most frequently rated as Q3 (58%) by Group 1. As was the case in Set 1, Stimuli 1 and 2 were usually perceived as Q2 (73% and 59% respectively), and stimulus 5 was judged as Q3 only 51% of the time. Group 2 only selected Q2 for all the stimuli in Set 2.

The results of Group 1 indicate that the optimal tonal contour for Q3 is characterized by an optimal length of the fall, see Fig. 3. The optimal fall was determined to be 13–33% of the V1 duration when the pitch range was 6 semitones, and 13% of the V1 duration when the pitch range was 3 semitones. A fall that was too long resulted in the lowest rate of Q3 answers (Q2 was perceived). However, in order for Q3 to be perceived there must be a perceivable fall period: sharp, almost 0 ms fall time resulted in even fewer Q3 answers.

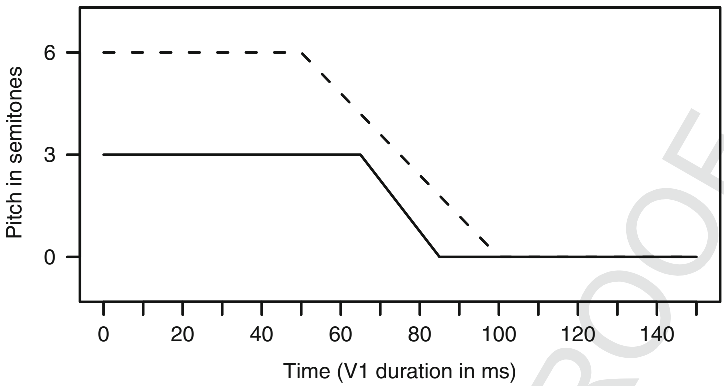


Fig. 3 The pitch contours of V1 of the stimuli with the most Q3 judgements. The *solid line* represents the pitch contour of Stimulus 4 with the pitch range of 3 st, the *dotted line* the pitch contour of Stimulus 3 with the pitch range of 6 st

3 Experiment 2

3.1 Materials and Methods

Using the same Q2 base word [sɑːtɑ] and re-synthesis method as in Experiment 1, two sets of five stimuli were created where the locus of the pitch fall was altered. The pitch fell during 50 ms (about 1/3 of the V1 duration). The start of the fall was varied by five 20 ms increments, resulting in a 10 ms high plateau and 90 ms low plateau in the first stimulus, and 90 ms high plateau and 10 ms low plateau in the last stimulus (see Fig. 4). In one set the F0 ranged from 100 Hz to 140 Hz (about 6 st; Set 3), in another from 100 Hz to 120 Hz (about 3 st; Set 4). The stimuli were presented to the test subjects in two blocks of 5 with 10 repetitions in random order (i.e. $2 \times 5 \times 10 = 100$ stimuli in total). The subjects were instructed to select whether a stimulus was a Q2 or Q3 word.

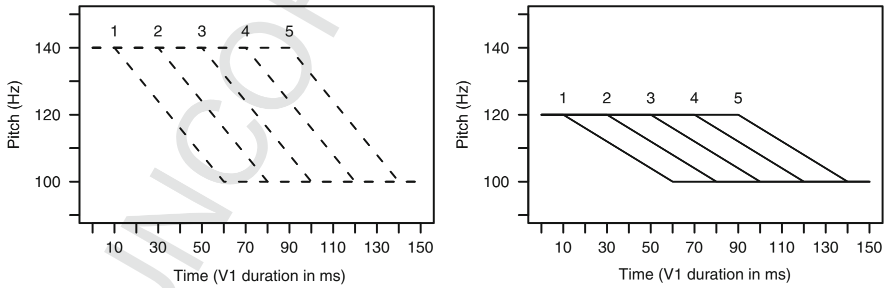


Fig. 4 Schematic pitch contours of the V1 of the stimuli. The starting point of the fall is marked with the stimulus number. *Left*: Set 3 with the pitch range from 100 Hz to 140 Hz. *Right*: Set 4 with the pitch range from 100 Hz to 120 Hz. In V2 the pitch continued at 100 Hz in all stimuli

The test subjects were the same 22 native Estonian speakers who participated in Experiment 1. The test setup and instructions were the same as in Experiment 1.

3.2 Results

The results of Experiment 2 (see Table 3 and Fig. 5) demonstrate that the alignment of the pitch contour significantly influenced Q2/Q3 perception in both groups. However, the difference between the main group and the deviating group remains.

For Set 3 with the pitch range of 6 st, participants most frequently selected Q3 for Stimuli 2 and 3, while for Stimuli 1 and 5 participants tended to select Q2. Stimuli 2 and 3 had the highest frequency of Q3 answers. The pitch contour in Stimulus 3 is identical to Stimulus 3 in Experiment 1, which received the highest Q3 response. In Stimulus 3, the duration of the high plateau, pitch fall, and low plateau were equal to 1/3 of the V1 duration. In Stimulus 2 the high plateau duration was 20% of the V1 duration, pitch fell during 33% of the V1 duration, and the low plateau duration was 47% of the V1 duration.

For Set 4 with the pitch range of 3 st all the stimuli received less Q3 responses than in Set 3. Once again Stimulus 3 received the most Q3 responses. Stimulus 1 which has the shortest high plateau and Stimulus 5 which has the shortest low plateau were perceived as Q2.

The results of Group 2 differ from the main group in the overall rate of Q3 responses. The stimuli in Experiment 2 were all perceived as Q2 by Group 2 in

Table 3 The percentage of Q3 responses in Experiment 2

	Stimulus 1	Stimulus 2	Stimulus 3	Stimulus 4	Stimulus 5
Set 3 Group 1	48%	72%	71%	56%	24%
Set 3 Group 2	20%	45%	42%	37%	22%
Set 4 Group 1	26%	53%	63%	43%	19%
Set 4 Group 2	10%	13%	20%	20%	5%

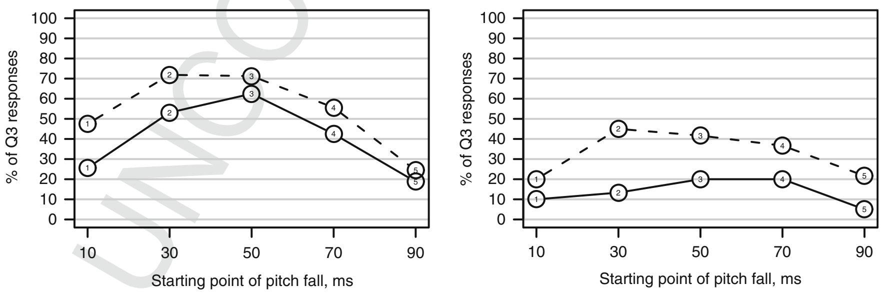


Fig. 5 Percentage of Q3 responses in Experiment 2. Set 3 is connected with *dotted lines*, Set 4 with *solid lines*. *Left*: the results of Group 1. *Right*: Results of Group 2

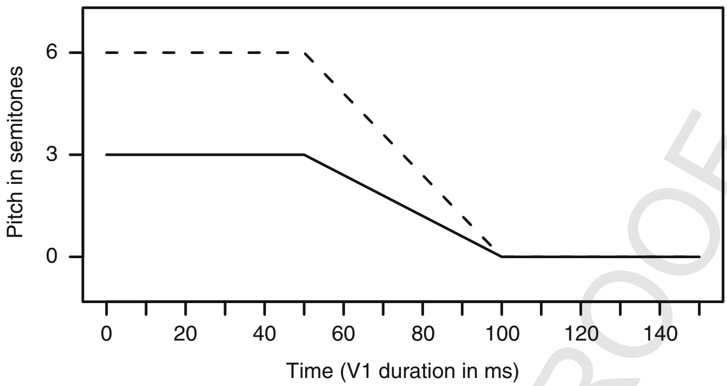


Fig. 6 The V1 pitch contours of the stimuli with the most Q3 judgements. The *solid line* represents the pitch contour of Stimulus 3 with the pitch range of 3 st, the *dotted line* the pitch contour of Stimulus 3 with the pitch range of 6 st

more than 50% of the cases, but the Q3 responses given by Group 2 follow the same pattern as the responses given by Group 1: Stimuli 2 and 3 were rated Q3 more frequently than Stimuli 1 and 5.

The results show that the difference between the high and low parts of the pitch contour, the location, and the duration of the falling period influence the perception of Q3. While the pitch range is larger, the fall can also be earlier. If the pitch fall was too early or late during V1, Q2 was perceived. Since the durational structure of the base word was not modified and was typical for Q2, the most important result is that pitch contour with high and low plateaus realized within V1 favours the perception of Q3. The pitch contours which elicited the most Q3 answers are presented in Fig. 6.

4 Discussion

Both experiments in this study show that it is possible to generate an overlong (Q3) word by changing the pitch contour of a long (Q2) word. A disyllabic Q2 word with typical duration ratios is perceived as a Q3 word if its 'level' pitch contour in the initial stressed syllable is changed to 'falling', following certain parameters.

Experiment 1 demonstrates that the optimal Q3 pitch contour comprises a sharp pitch fall and sufficiently long high and low plateaus. However, the perception of Q3 is disturbed if the duration of the fall is too short. The rate of the Q3 responses was the highest when the fall constituted 13–33% and the high and the low plateau both 33–43% of the V1 duration.

A group of 6 subjects gave unexpected responses to the stimuli of Experiment 1. They perceived Q3 overwhelmingly when there was a large pitch range (6 st) and Q2 when there was a smaller pitch range (3 st). No social or regional differences from Group 1 in the background of the subjects were found. As reported by Lippus and Pajusalu (2009), the subjects from the western and central Estonian dialect areas tend to be more influenced by the pitch cue when perceiving quantity than the subjects from the southern and eastern Estonian dialect areas. In this case subjects in both groups come mainly from central Estonian areas. Moreover, subjects in Group 2 were influenced by the pitch because otherwise they would have heard only Q2 regardless of the pitch pattern, as the temporal structure of the Q2 base word was not modified.

The results of Group 2 in Experiment 1 could be explained as an effect of the adaptation level (cf. Helson 1947), i.e. if several features of the stimuli – the turning point of the pitch, durations of the high and low plateaus and the length of the fall – were simultaneously altered, then these subjects focused only on the pitch range between high and low plateaus. This result indicates that a falling pitch contour with a significant pitch range is the primary pitch feature for distinguishing Q3 words.

According to the results of Experiment 2, the placement of the turning point of the pitch contour is important for the perception of Q3. Q2 is usually perceived if the turning point of the pitch is in the very beginning of the V1 or in the second half of the V1, but if the turning point is at the distance of 20–40% from the beginning of the V1, Q3 is most frequently perceived. In the case of the pitch range of 3 semitones, the rate of Q3 responses was highest when the high plateau, pitch fall and low plateau were all 33% of the vowel duration. The turning point can also be earlier, if the pitch range is great (6 st), leaving the high plateau as 20% of the vowel duration, and the low plateau as 47%. These results indicate that both the high plateau and the low plateau have to be sufficiently long in order to perceive Q3. However, there is still a valid restriction concerning the turning point; it must be located in the first half of the V1.

Group 2, who in Experiment 1 had difficulties in the perception of several interrelated factors, gave different responses in Experiment 2 if the duration of the fall was fixed. To some extent their results in Experiment 2 are similar to the responses given by Group 1. None of the stimuli in Experiment 2 were perceived as Q3 by Group 2, but to some extent the pitch contours that caused perception of Q3 for Group 1 also predisposed perception of Q3 for the deviating group. Thus, it is reasonable to state that after a sufficiently great pitch range a balanced duration of high and low plateaus is the second important feature for perceiving Q3.

Two main features – a significant pitch range and an optimal length of both high and low plateaus – are essential for the perception of Q3 pitch contour. The required sharpness of the fall is ultimately determined by the realization of these features. However, as the results of Experiment 1 show, the fall cannot be too short. Apparently the fall has its own status in the formation of the optimal pitch contour for Q3.

5 Conclusion

The results of the study largely confirm earlier statements about the role of the pitch contour in the perception of Estonian quantity degrees. Previous research has shown that a conflicting combination of pitch and temporal cues can disturb the identification of the quantity. Our results further demonstrate that words with an identical durational structure can be differently perceived depending on their pitch contour. Our results additionally help to specify the importance of various factors in the creation of an optimal Q3 pitch contour. The three main factors which mark the most characteristic falling contour of Estonian Q3 words are a significant pitch range between the high part and the following low part of the pitch, a sufficiently long high and low plateaus and a relatively sharp fall.

Acknowledgments We would like to thank all our test subjects and Einar Meister for helping to find the test subjects. We are also very grateful to Eva Liina Asu-Garcia, and to anonymous reviewers for their comments on this paper, and to Cameron Robert Rule for editing the language of this paper. The present research was partly supported by the Estonian Science Foundation grant No. 7904.

References

Asu, Eva Liina. 2004. *The Phonetics and Phonology of Estonian Intonation*. Doctoral dissertation. University of Cambridge.

Asu, Eva Liina and Francis Nolan. 1999. The effect of intonation on pitch cues to the Estonian quantity contrast. In *Proceedings of the 14th International Congress of Phonetic Sciences, San Francisco, USA, 1–7 August 1999*, Vol. 3, 1873–1876. San Francisco.

Asu, Eva Liina and Francis Nolan. 2007. The analysis of low accentuation in Estonian. *Language and Speech* 50(4): 567–588.

Asu, Eva Liina, Pärtel Lippus, Pire Teras and Tuuli Tuisk. 2009. The realization of Estonian quantity characteristics in spontaneous speech. In Martti Vainio, Reijo Aulanko and Olli Aaltonen, (eds.) *Nordic Prosody. Proceedings of the Xth Conference, Helsinki 2008*, 49–56. Frankfurt: Peter Lang.

Boersma, Paul and David Weenink. 2007. *Praat: Doing Phonetics by Computer (Version 4.6.31)* [Computer program]. http://www.praat.org/. Retrieved 12 October 2007.

Eek, Arvo. 1980a. Estonian quantity: notes on the perception of duration. In Arvo Eek, (ed.) *Estonian Papers in Phonetics 1979*, 5–29. Tallinn: Academy of Sciences of the Estonian S. S. R. Institute of Language and Literature.

Eek, Arvo. 1980b. Further information on the perception of Estonian quantity. In Arvo Eek, (ed.) *Estonian Papers in Phonetics 1979*, 31–56. Tallinn: Academy of Sciences of the Estonian S. S. R. Institute of Language and Literature.

Eek, Arvo and Einar Meister. 2003. Foneetilisi katseid ja arutlusi kvantiteedi alalt (I). Häälikukestusi muutvad kontekstid ja välde. *Keel ja Kirjandus* 11–12: 815–837, 904–918.

Eek, Arvo and Einar Meister. 2004. Foneetilisi katseid ja arutlusi kvantiteedi alalt (II). Takt, silp ja välde. *Keel ja Kirjandus* 4–5: 251–271, 336–357.

Helson, H. 1947. Adaptation-level as frame of reference for prediction of psychophysical data. *American Journal of Psychology* 60: 1–29.

Kask, Arnold. 1972. *Eesti keele ajalooline grammatika*. Tartu: Tartu Riiklik Ülikool.

Lehiste, Ilse. 1960. Segmental and syllabic quantity in Estonian. In *American Studies in Uralic Linguistics* 1, 21–82. Bloomington.

Lehiste, Ilse. 1970–1975. Experiments with synthetic speech concerning quantity in Estonian. In Valmen Hallap (ed.) *Congressus Tertius Internationalis Fenno-Ugristarum, Tallinae habitus, 17–23. VIII 1970. Pars I: Acta Linguistica*., 254–69. Tallinn: Valgus.

Lehiste, Ilse. 1976. Influence of fundamental frequency pattern on the perception of duration. *Journal of Phonetics* 4: 113–117.

Lehiste, Ilse. 1978. Polytonicity in the area surrounding the Baltic Sea. In Gårding, E., Bruce, G., Bannert, R, (eds.) *Nordic Prosody: Papers from a Symposium*, 237–247. Lund: Department of Linguistics, Lund University.

Lehiste Ilse. 1997. Search for phonetic correlates in Estonian Prosody. In Ilse Lehiste and Jaan Ross, (eds.) *Estonian Prosody: Papers from a Symposium*, 11–35. Tallinn: Institute of Estonian Language.

Lehiste, Ilse. 2003. Prosodic change in progress: from quantity language to accent language. In Paula Fikkert and Haike Jacobs, (eds.) *Development in Prosodic Systems*, 47–66. Berlin, New York: Mouton de Gruyter.

Lehiste Ilse and Douglas G. Danforth. 1977. Foneettisten vihjeiden hierarkia viron kvantiteetin havaitsemisessa. *Virittäjä* 4: 404–411.

Liiv, Georg. 1961. Eesti keele kolme vältusastme vokaalide kestus ja meloodiatüübid. *Keel ja Kirjandus* 7, 8, 412–424, 480–490.

Lippus, Pärtel and Karl Pajusalu. 2009. Regional variation in the perception of Estonian quantity. In Martti Vainio, Reijo Aulanko and Olli Aaltonen, (eds.) *Nordic Prosody. Proceedings of the Xth Conference, Helsinki 2008*, 151–157. Frankfurt: Peter Lang.

Lippus, Pärtel, Karl Pajusalu and Jüri Allik. 2007. The Tonal Component in Perception of the Estonian Quantity. In Jürgen Trouvain and William J. Barry, (eds.) *The Proceedings of the 16th International Congress of Phonetic Sciences: 16th International Congress of Phonetic Sciences, Saarbrücken, Germany, 6–10 August 2007*, 1049–1052. Saarbrücken.

Lippus, Pärtel, Karl Pajusalu and Jüri Allik. 2009. The tonal component of Estonian quantity in native and non-native perception. *Journal of Phonetics* 37: 388–396.

Nolan, Francis and Eva Liina Asu. 2009. The pairwise variability index and coexisting rhytms in language. *Phonetica* 66: 64–77.

Remmel, Mart. 1975. *The Phonetic Scope of Estonian: Some Specifications*. Preprint KKI-5. Tallinn: Academy of Sciences of the Estonian S.S.R. Institute of Language and Literature.

Traunmüller, Hartmund and Diana Krull. 2003. The effect of local speaking rate on the perception of quantity in Estonian. *Phonetica* 60: 187–207.

Viitso, Tiit-Rein. 2003. Phonology, morphology and word formation. In Mati Erelt (ed.) *Estonian Language*, 9–92. Tallinn: Estonian Academy Publishers.

All Depressors are Not Alike: A Comparison of Shanghai Chinese and Zulu*

Yiya Chen and Laura J. Downing

1 Introduction

It is cross-linguistically rather common for voiced consonants to have a lowering effect on tone realization (Hombert 1978; Bradshaw 1999; Tang 2008; Lee 2008). More surprising are languages where (some) voiceless consonants also have a pitch lowering effect. In this paper, we take a closer look at two such languages, Shanghai Chinese and Zulu. We have chosen to compare these two languages in order to follow up on a recent proposal in Jessen and Roux (2002) that the same [slack voice] feature which Ladefoged and Maddieson (1996) suggests characterizes the voiceless depressor consonants of Shanghai Chinese also characterizes Nguni depressors. (The Nguni Bantu languages include Ndebele, Phuthi, Swati, Xhosa and Zulu.) The basis for comparing these two languages is that both have what is described as the same three voiceless stop series: voiceless aspirated, voiceless unaspirated, and voiceless depressor. Further, Jessen and Roux suggest that there is parallel phonetic implementation of the [slack voice] feature in Nguni languages and Shanghai. Since Nguni depressors (like Shanghai Chinese depressors) are, in fact, voiceless, they further propose that *f0* lowering following the depressor consonants 'compensates' for the lack of phonetic voicing during stop closure. Our study pursues the comparison between Zulu and Shanghai Chinese depressors in detail.

While we agree with Jessen and Roux that the same [slack voice] feature can be used to characterize depressors in Shanghai Chinese and Nguni

L. Downing (✉)
ZAS (Berlin), Berlin, Germany
e-mail: downing@zas.gwz-berlin.de

*We would like to thank our Zulu language consultants for their help in constructing the Zulu data sets and for their patience in making the recordings on which the phonetic analysis is based. An earlier version of this paper was presented at the TIE3 conference in Lisbon. We thank the audience of that conference, along with four anonymous reviewers and the editors of this volume, for helpful comments.

S. Frota et al. (eds.), *Prosodic Categories: Production, Perception and Comprehension*, Studies in Natural Language and Linguistic Theory 82,
DOI 10.1007/978-94-007-0137-3_11, © Springer Science+Business Media B.V. 2011

languages, we disagree with their proposal that the two languages implement this feature in a parallel way. We also argue that *f0* lowering does not compensate for lack of phonetic voicing. Instead, following work like Kingston and Diehl (1994) we propose that the phonetic interpretation of features like [slack voice] is subject to language-specific variation, conditioned in Zulu and Shanghai Chinese, we show, by differences in the phonology of the tone system of the two languages.

The paper is organized as follows. Section 2 presents our production study of tone-segment interactions in Zulu, in comparison with the Shanghai Chinese data reported in Chen (2007, submitted). Section 3 discusses the implications of our results for Jessen and Roux's (2002) proposal that these two languages implement the same [slack voice] depressor feature in a parallel way. We will show that there are, in fact, important differences in the phonetics of depressors in the two languages. In Section 4, we show that the variation in the implementation of depression is conditioned by the phonology, not simply an automatic result of parallel implementation of the same phonological phonetic feature. In Section 5, we justify our choice of [slack voice] to characterize depressor consonants in the two languages and take up the issue of how much phonetic variation can be associated with a particular phonological feature. We conclude in Section 6.

2 Production Study

The goal of this study is to investigate the following two related questions which arise as a natural follow-up to Jessen and Roux (2002). The first question is whether pitch depression is implemented on the vowel following the target consonant in the same way in Zulu and Shanghai Chinese, as they suggest. To this end, we examined the specific pattern of *f0* depression in Zulu and compared that to similar Shanghai Chinese data, reported in Chen (2007, submitted). Specifically, we examined whether *f0* lowering is a consistent correlate of voiceless depressor consonants in both languages, and if so, whether the domain of *f0* lowering (e.g., the beginning portion of the target syllable or the whole syllable) is comparable in the two languages.

The second question we investigated concerns the relation between *f0* lowering and phonetic voicing during stop closure in the two languages. We examined whether *f0* lowering and phonetic voicing 'trade off' in the same way during depressor stop closure in both languages. We also investigated a related question, namely, whether the stops in the two languages that show phonetic voicing – implosives in Zulu and word-medial depressors in Shanghai Chinese – display a pattern of *f0* lowering comparable to the effect of voiceless depressors in Zulu.

2.1 Methods

2.1.1 Stimuli

The Zulu data set was constructed to be optimally comparable to the data sets used in existing studies of Shanghai Chinese (Chen 2007, submitted) and Xhosa (Jessen and Roux 2002), without compromising the linguistic characteristics of the target language. Two main factors were controlled for: the tonal context and the morphosyntactic position of the target syllable. We discuss each of these in turn.

In Shanghai Chinese, tonal contrasts over the prosodic-word medial syllables are neutralized: these medial syllables surface with *f0* contours that are determined by the preceding lexical tones. (We discuss this pattern further in Section 4.1, below.) Chen (2007, submitted) reports that the specific pattern of consonant-induced *f0* differences over the medial syllables also varies according to the preceding tonal context. Jessen and Roux (2002) did not mention any contextual tonal variation in Xhosa; they included only target syllables with Low tone and no specific information about the tone of the preceding syllable was provided.

In Zulu, there is contrastive High vs. Low (default) tone. Prior studies on other tone languages suggest that tonal realization is influenced by its context, with preceding lexical tones usually exerting significant influence over the following tone (Gandour et al. 1994; Xu 1997). Since it was not clear whether and how lexical tones coarticulate in Zulu, we controlled the preceding lexical tone in addition to the tone of the target syllable. This improves on the data collected by Jessen and Roux (2002) for Xhosa and also makes our data set more comparable to the one collected by Chen (2007, submitted) for Shanghai Chinese. The tonal contexts included in the Zulu data set are shown in Table 1:

Another factor we controlled for is the morphosyntactic position of the target syllable. Like the Xhosa data in Jessen and Roux, the Zulu data set was comprised of disyllabic (verb) stems, preceded by the inflectional prefixes (i.e., subject agreement prefix and tense/focus prefix) required to allow the verb stem to be pronounced in isolation.[1] The subject prefixes contrast for tone, and so allow for the stem-initial stop to be preceded by either High or Low toned syllables. The target syllable was the stem-initial stressed syllable, which also

Table 1 Tonal context of target syllables for Zulu: four different tonal contexts for each voiceless stop type: preceding syllable (High/Low) + target syllable (High/Low)

Tone Context: Preceding + **Target**	Jessen and Roux (2002)	Current study
High + **High**	-	√
Low + **High**	-	√
High + **Low**	√(Mixed data)	√
Low + **Low**		√

[1]See work like Schadeberg (2003) for discussion of the morphological structure of the Bantu verb stem.

contrasts for High vs. Low tone. The data in (1), below, illustrates these tonal and morphosyntactic contexts for Zulu aspirated stops.[2] Note that the disyllabic stem is preceded by ‘ = ’, and the target syllable is bolded.

(1)

Tone context	Zulu verbs	Gloss
(prefix) High + (stem) **High**	bá-yá = **tháá**nda	‘they like’
(prefix) Low + (stem) **High**	si-ya = **tháá**nda	‘we like’
(prefix) High + (stem) **Low**	bá-yá = **phaa**nda	‘they are digging’
(prefix) Low + (stem) **Low**	si-ya = **phaa**nda	‘we are digging’

Table 2 illustrates the Zulu and the Shanghai Chinese data sets with one word for each of the voiceless stops in the tonal context of High followed by a Low tone. Note that the whole data set in Shanghai Chinese data in Chen (2007, submitted) were composed of bi-syllabic names (*XY*) with the family name *X* carrying three different lexical tones (Falling [h-HL]; high-register Rising [h-LH] and low-register Rising [l-LH]). The second syllable *Y* (i.e. the part as the given name) varied in terms of their laryngeal contrast: voiceless unaspirated, voiceless aspirated, and depressor. These disyllabic names form tone sandhi domains (to be discussed in more detail in Section 4.1) and were elicited in the subject position of different carrier sentences. Due to tone sandhi, the Falling tone [h-HL] is realized with a high F0 value over the first syllable while the second syllable ends with a rather low F0 value, and is therefore comparable to the High - Low tone context in Zulu (i.e., the H_X_L tonal context where X refers to the onset consonant of the second syllable).

2.1.2 Discourse Context

It is well known that the *f0* realization of lexical tones is influenced significantly by the discourse context in which a sentence is uttered (e.g., Chen and

Table 2 Zulu and Shanghai Chinese partial data set: voiceless stops in the H_X_L tonal context – target syllable is underlined and low tone unmarked

Stop Type	Zulu	Shanghai
Aspirated	bá-yá = khaaba ‘they are kicking’	[zʼɑŋtʰɪŋ] ‘name’
Unaspirated	bá-yá = kaakwa ‘they are being surrounded’	[zʼɑŋtɪŋ] ‘name’
Depressor	bá-yá = bheeka ‘they are watching’	[zʼɑŋdɪŋ] ‘name’

[2]The complete word list analyzed for the Zulu study reported on here is found in Appendix 1. We thank our colleague, Leston Buell, a Zulu specialist (Buell 2005), for his help in constructing the Zulu word list. The Zulu words in this chapter are cited in the orthography, except that penult stress (vowel lengthening) is indicated and accents indicate High tone (Low tone is not marked). See Appendix 2 for a complete list of Zulu consonants with their phonetic description. See Section. 4.2, below, for a sketch of the Zulu tone system.

Gussenhoven 2008). The stimulus sentences were therefore elicited in a controlled discourse context (hereafter referred to as the Focus condition): the target word is in new information focus, elicited with a WH-question on the target word, as illustrated below:

(2)

Question:	Sénzaa-ni?	'What are you doing?'
Answer:	Si-ya-péénda.	'We are painting.'

This discourse context again was chosen to make the Zulu data optimally comparable to the Shanghai Chinese data analyzed in Chen (2007, submitted), where the target words were elicited in the Focus condition as well as the No Focus condition.[3]

2.1.3 Participants and Recording Procedure

Two female Zulu speakers participated in the experiment, one in her twenties and the other in her fifties at the time of recording. According to their self-report, both speakers are native speakers of Zulu. The younger speaker was living temporarily in the Netherlands, while the older one had immigrated to the Netherlands some years before but speaks Zulu regularly with family members in South Africa.

Before the recording, each participant was presented with a printed version of the word list, and we went through the list with the participants. The purpose of this procedure was to ensure that both speakers knew the words on the list and considered them not uncommon in daily usage. Our original word list contained more than 30 items. In talking to the speakers before the recording, we learned that some words that the elder speaker considered common were not necessarily so for the younger one. Speakers were assured that they did not have to produce these uncommon words. Both speakers also suggested additional words that were segmentally very comparable, and we added these items to the wordlist. It is important to note, however, that the majority of the items on our original word list were found to be good, familiar Zulu words by both our speakers.

Both participants were recorded in a soundproof booth at the Phonetics Lab at the Leiden University Center of Linguistics, via Adobe Audition. The stimuli were presented in Internet Explorer via a JAVA program, which randomized the order of the stimuli each time the program was run, except for the first and last two filler items. The program presented the target sentences one at a time. During the recording, subjects were first given the sentences in Zulu orthography on a computer screen. They were then played a pre-recorded oral question, read by one of the speakers on a different occasion. The data were digitized at a

[3]In the No Focus condition, the target words were produced as old information, elicited with a WH-question on a constituent after the target word, later in the sentence.

44100 Hz sampling rate with 16-bit resolution and later downsampled to 22050 Hz in GoldWave before acoustic analyses. We analyzed two repetitions of the same set of tokens from each speaker.

2.1.4 Data Analyses

The data analyzed and presented below were based on part of the corpus collected. Since we did not have access to a large number of speakers of Zulu, we chose only words which are as comparable as possible, so that we could gain a good understanding of the contribution of the variables we controlled for (i.e. consonant onset and tonal contexts) without being distracted by variations that are outside the interests of this study. Appendix 1 lists the words chosen for analysis.

Two aspects of the data were analyzed: (1) voicing of the target consonant closure and (2) *f0* over the target syllables, measured from the first regular vocal pulse. Occurrence of voicing was ascertained via eyeball inspection of spectrograms. As for *f0*, we first marked the beginning and end of the vowel of the target syllable in Praat (Boersma and Weenink 1996–2001), based on the periodicity in the acoustic waveform and supplemented by spectrographic analyses. Since our goal was to understand the effect of consonant onset on the *f0* of its following vowel, the beginning of the vowel was identified as the onset of the first clear periodic pattern in the acoustic signal. The *f0* of 20 points at proportionally equal time intervals between the start and offset of the vowel was measured. To plot the *f0* contours, these values were averaged across repetitions and items in the same tonal context.

2.2 Results

Figures 1 and 2 show that the depressor consonants have a consistent *f0* lowering effect in both languages. The specific patterns of *f0* lowering, however, differ in the two languages. As we can see in Fig. 1, in Zulu, the effect of the depressor on *f0* remains salient during most of the target syllable, except in the Low-Low tonal context. (The Zulu data set used to elicit the stops in the different tonal contexts is given in Appendix 1.)

This contrasts with what we find in the Shanghai data. As we can see in Fig. 2, the effect of the different consonants on *f0* wanes much faster during the time course of the target syllable. (The L_L combination is not found in Shanghai.)

We also found a difference between the two languages with regard to the relation between *f0* lowering and phonetic voicing. In Zulu, we found no closure voicing on the depressors at all, in any position, confirming earlier phonetic studies, which consistently affirm the voiceless nature of the depressors (Doke 1961; Traill et al. 1987; Giannini et al. 1988). However, the voiceless depressors consistently have a lowering effect on *f0*. In contrast, the implosive stops of Zulu, which are phonetically voiced, show only a small *f0* lowering effect, if any,

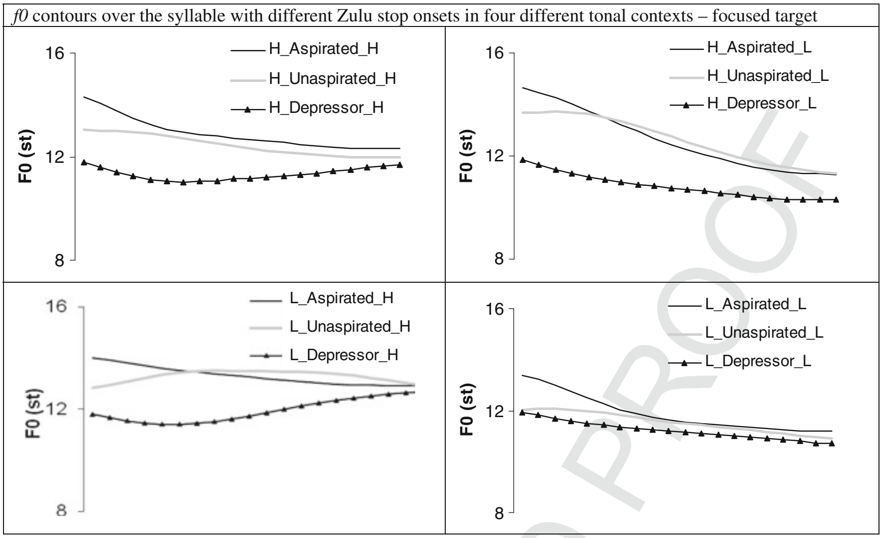


Fig. 1 Effect of aspirated, unaspirated (or ejective) and depressor consonants on pitch of the target syllable in Zulu

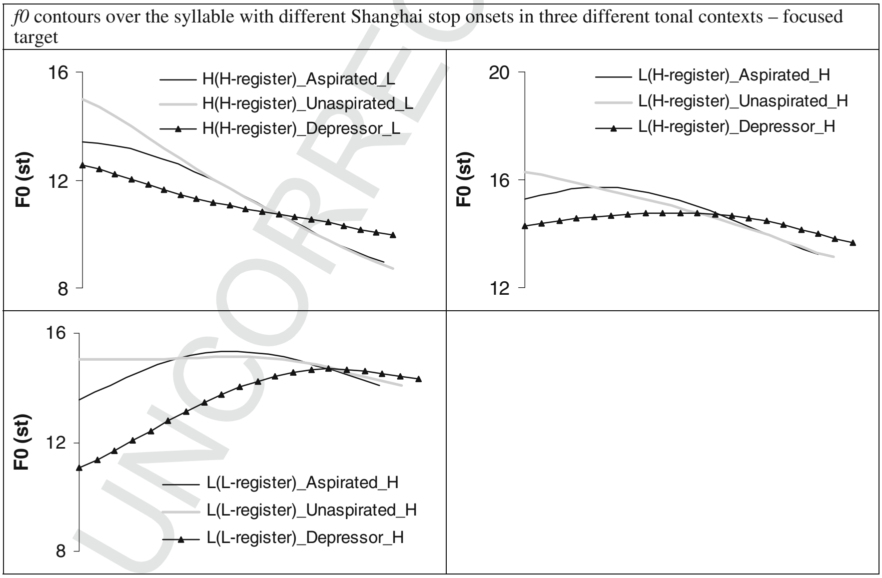


Fig. 2 Effect of aspirated, unaspirated and depressor consonants on pitch of the target syllable in Shanghai (data replotted from Chen 2007)

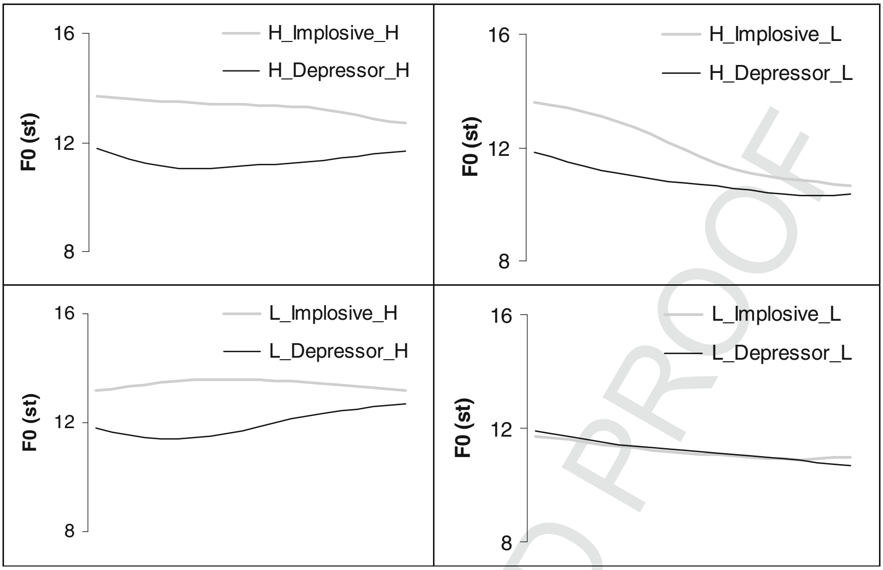


Fig. 3 Effect of implosive vs. depressor consonants on pitch of the target syllable in Zulu

compared to the voiceless depressors. This is shown in Figure 3. (See Appendix 1 for the Zulu data set used to elicit the stops in the different tonal contexts.)

The relation between phonetic voicing and *f0* lowering is somewhat different in Shanghai Chinese. Depressor consonants always have an *f0* lowering effect, whether they are in word-initial position,[4] where they are voiceless, or in word-medial position, where they are consistently voiced (Chen 2007, submitted).[5] However, in word-medial position, where they are voiced, *f0* lowering is apparently not phonologized since a High tone can spread from the preceding syllable; it is not prohibited in this position.

3 Discussion

In the light of this phonetic study and other work cited below, we now evaluate three claims Jessen and Roux (2002) make about the relation between *f0* lowering and depressor consonants in Nguni languages and Shanghai Chinese and show there are problems with each of these claims. As a background to this critique, it is

[4]In word-initial position, the voiceless depressor can only occur with low register Rising tones. In other words, the f0 lowering effect has been phonologized.

[5]Chen (2007, submitted) observes that in the No Focus context, the depressor lowering effect is much less salient. See Section 4.1, below, for more detailed discussion of how tone sandhi processes interact with the phonetic aspects of tone realization described here.

useful to summarize briefly the main points of Jessen and Roux. This study, the most recent one we know of the phonetic properties of Nguni depressors, investigates the influence of Xhosa stops – in particular, the voiceless depressors – on the *f0* of the following vowel. Xhosa is an Nguni Bantu language closely related to Zulu, and the phonology of depressor consonants in Xhosa is essentially identical to that of Zulu, as work like Cassimjee (1998) and Cassimjee and Kisseberth (1998, 2001) shows. (See Section. 4.2, below, for a sketch of the phonology of depressors in Zulu.) Jessen and Roux's results agree, in many of their essentials, with earlier phonetic studies of depressor consonants in other Nguni languages (Zulu and Swati), like Traill et al. (1987), Traill (1990), Giannini et al. (1988), Wright (1992), Maddieson (2003) and Strazny (2003). In Xhosa, as in these other Nguni languages, the depressor stop consonants are voiceless during closure and have a significant lowering effect on the *f0* of the following vowel.[6] Our study also confirms these findings for Zulu. In Xhosa, as in these other Nguni languages, breathy voice is not a systematic accompaniment of the depressors (Jessen and Roux: 37), though it does variably occur. There are some important differences, too, between their study and previous studies on other Nguni languages. First, they only investigated the influence of depressors (and other stops) on vowels realized with a Low tone. All other studies, like ours, have investigated the influence of depressors (and other stops) on both High-toned and Low-toned vowels. Further, Jessen and Roux (2002: 40) report that in Xhosa, implosives are accompanied by *f0* depression 'similar to that found after voiced stops [i.e., depressors]'. The studies of other Nguni languages cited above, like ours, have not found that the implosives lower pitch in the same way that depressors do. (See Figure 3.) We return to these differences in the discussion below.

We begin our critique with the concluding claim of Jessen and Roux's paper, namely, that the feature [slack voice] characterizes depressor consonants in Nguni languages (like Xhosa and Zulu). This proposal is not entirely new: work like Khumalo (1981) and Traill et al. (1987) also suggest that a [slack voice] feature is involved in Zulu depression. In the case of Traill et al., this proposal is supported by a laryngoscopic study. What is new is that Jessen and Roux draw an explicit analogy with Shanghai Chinese, and propose that the same [slack voice] feature that Ladefoged and Maddieson (1996) suggest characterizes the voiceless depressors of Shanghai Chinese is also found in Nguni depressors. One problem with this claim is that while a [slack voice] feature may be involved in both languages, Traill et al.'s (1987) laryngoscopic study of Zulu establishes that a devoicing gesture is essential to explain the lack of voicing in all contexts. That is, the [slack voice] feature alone does not characterize the phonetic laryngeal properties of Zulu depressors. Jessen and Roux, however, do not provide a concrete explanation for this finding. Another problem with this

[6]Confusingly, Jessen & Roux (2002) refer to the depressors as 'voiced' throughout their paper, in contradiction to their finding that, in all possible contexts, they are phonetically not voiced.

claim, which we take up in detail presently, is that there are important differences in the phonetic implementation of consonant 'depression' in Zulu compared to Shanghai Chinese.

Since Nguni depressor stops are, in fact, voiceless during closure, [slack voice] would not seem to be an obvious choice to define this class of consonants. Jessen and Roux (2002: 39) claim, though, that the depressor effect in Nguni could be understood as akin to historical tonogenesis in Southeast Asian languages, including Shanghai Chinese (Matisoff 1973; Cao and Maddieson 1992; Svantesson and House 2006; Brunelle 2008): *f0* lowering 'compensates' for lack of (or loss of) phonetic voicing during closure. There are several problems with this claim. For one thing, it assumes that phonetic voicing and *f0* lowering originally correlate with each other. There are, however, numerous exceptions to this correlation, where *f0* lowering correlates with phonetically voiceless consonants and phonetically voiced consonants do not correlate with *f0* lowering. Indeed, the voiceless depressors of Shanghai Chinese and Nguni Bantu are not the only types of voiceless consonants associated with *f0* lowering. As work like Chen (submitted), Downing and Gick (2005), Lee (2008: 57) and Tang (2008: 25-26) show, aspirated voiceless stops and voiceless fricatives, especially those with long duration frication noise, and glottalized voiceless consonants commonly correlate with Low tone or *f0* lowering. It is a mistake to automatically link low pitch or Low tone only with voicing. And, as shown in Fig. 3, implosives in Zulu are voiced but in most contexts they are realized with a raised *f0*, more comparable to the aspirated stops than the depressors. Indeed, as Tang's (2008) recent survey of consonant-tone interactions shows, implosives have a variable effect on tone cross-linguistically, raising pitch in some languages and lowering it in others. While Jessen and Roux (2002: 40) report that in Xhosa, *f0* depression is 'similar to that found after voiced stops [i.e., depressors] with those speakers that produced fully voiced implosives,' this result is probably due to the limited tonal contexts they investigated. As we can see in Fig. 3, above, the only tonal context where implosives induce tone lowering similar to that of depressors is when the consonants are preceded and followed by Low tones: i.e., in the only tonal context that Jessen and Roux investigated. This explains why Jessen and Roux reach a different conclusion about the influence of implosives on *f0* than other studies of Nguni tone-segment interactions. In other tonal contexts, implosives clearly do not have a pitch lowering effect comparable to the depressors.

Another problem with assuming that *f0* lowering 'compensates' for phonetic voicing, akin to tonogenesis in other languages, is that in Zulu phonetic voicing is not observed in the depressor stop consonants and also has never been documented.[7] Nevertheless, *f0* lowering is consistently observed. Instead, *f0*

[7]See Downing (2009) for a detailed review of phonetic studies of Nguni depressor consonants. While work like Bradshaw (1999) and Clements (2003) suggests that Nguni depressor stops were historically voiced, Schadeberg's (2009) detailed discussion of the historical source of these consonants demonstrates that there is no empirical basis for this claim.

lowering in Zulu may well serve as a cue to signal some other laryngeal contrast, rather than voicing. Traill et al. (1987) show that depressor stops in Zulu are realized with slack voice plus a devoicing gesture. Perhaps *f0* lowering enhances or correlates with this combination of laryngeal configurations, rather than with slack voice alone.

A final problem with assuming that *f0* lowering compensates for phonetic voicing is that a balanced trade-off relation between voicing and tone is not systematically found in data from other languages, like Korean (Jun 1996; Silva 2006) or Shanghai Chinese, that have three voiceless stop series. In Shanghai Chinese – our main comparison language – what we find is that in the initial position *f0* register contrasts are phonologized (Xu and Tang 1988; Zee and Maddieson 1980). Even though all three stops are voiceless in this position, depressor consonants are followed by syllables with a low register tone, and other voiceless stops are followed by syllables with a high register tone. As Chen (2007) as well as other phonetic studies (e.g., Cao and Maddieson 1992, Ren 1992) show, in medial position, we do find consistent phonetic voicing on the depressors, however. Strikingly, the presence of voicing in medial position nevertheless introduces pitch lowering following these stops, which would not be expected if there is indeed a balanced trade-off relation between voicing and Low tone or low tone register. This clearly argues against a straightforward trade off relation between voicing and *f0* lowering.

A third claim that Jessen and Roux (2002: 41-42) make is that the [slack voice] feature is motivated for Nguni depressors because we find parallels in the phonetic implementation of Nguni depressors and the voiceless depressors of Shanghai Chinese, which Ladefoged and Maddieson (1996) have characterized as [slack voice]. However, as we have already pointed out in Section 2, a careful look at the phonetic properties of depressor consonants in the two languages instead reveals important differences in phonetic implementation. First, the two languages differ in the temporal domain of the depressor effect. This can be seen in comparing Fig. 1 (Zulu) and 2 (Shanghai Chinese). In Zulu, the depressor effect stays longer in the syllable in medial position. In Shanghai Chinese, the depressor effect in medial position wanes earlier in the syllable. They also differ in breathiness. In Shanghai Chinese, breathiness has been observed on the vowel following the voiceless depressors in several studies (Cao and Maddieson 1992; Ren 1992).[8] In Zulu, no breathiness was observed in our study, confirming the findings of other phonetic studies of Zulu depressors like Traill et al. (1987) and Giannini et al. (1988). Further, they differ in phonetic voicing. In Shanghai Chinese, depressors are voiceless in initial position, but voiced in medial position. In Zulu, the depressor stops are not voiced whatever their position in the word, not even in medial (intervocalic) position. Finally, they differ in their effect on the *f0* of a following vowel. In Shanghai Chinese,

[8]See, too, Clements & Khatiwada (2007), which suggests that breathy voice is an expected cross-linguistic correlate of pitch lowering on a vowel following a phonetically voiceless consonant.

depressors do not systematically have the same lowering effect on *f0*: position in the word, focus and tonal context strongly condition the lowering effect (for details, see Chen, submitted). In Zulu, depressors systematically affect *f0* regardless of position and in all tonal contexts except the Low-Low context. These differences are hard to explain if depressor effects in both languages are to be attributed to the same [slack voice] laryngeal feature, with parallel phonetic implementation, as proposed by Jessen and Roux (2002).

In sum, it is hard to agree with Jessen and Roux's (2002) proposal that voiceless depressors in both Nguni languages like Zulu and in Shanghai can be simply characterized as having a [slack voice] feature, if the principal motivation for this choice is that [slack voice] is phonetically implemented in the same way, with lack of phonetic voicing in both languages similarly 'compensated' for by *f0* lowering.

4 Phonetic Implementation of Tonal Depression is Controlled by Phonology

In this section we develop an alternative proposal. We argue that we can use the same [slack voice] feature for depressor consonants in both languages, if we adopt Kingston and Diehl's (1994) proposal that the specific patterns of phonetic implementation of phonological features are controlled by phonology and are thus expected to vary from language to language.[9] We show that differences in the tonal phonologies of Shanghai and Nguni languages like Zulu can explain many of the observed differences in the effect of depressor consonants on *f0*. That is, the differences in the phonetic implementation of [slack voice] follow from differences in the tonal phonology of the two languages. To see this, we must briefly sketch the tone system of each language, beginning with Shanghai Chinese.

4.1 Sketch of the Tone System of Shanghai Chinese

In this section we sketch the essentials of the Shanghai Chinese tone system which are relevant to understanding the interaction of depressor consonants

[9] Confusingly, Jessen & Roux (2002) make contradictory proposals about how similar the phonetic implementation of a phonological feature must be in different languages. On the one hand, they point to parallels in the phonetic implementation of [slack voice] in Xhosa and Shanghai Chinese to motivate the choice of this feature to characterize Xhosa depressors. On the other hand, they suggest (p. 39), following Kingston & Diehl (1994), that features like [voice] can have different phonetic implementations in different languages, and Xhosa could be considered a language where only a low-level feature, *f0* lowering, implements [voice]. Since we are interested in pursuing their suggestion that Nguni depressors show parallels with Shanghai Chinese depressors, we have also assumed the strictest possible interpretation of their first proposal, namely, that the two languages implement the same feature in a parallel way. We return to this point in Section 5, below.

with tone.[10] In Shanghai Chinese, as in other Chinese languages, tones are distributed over two tonal registers. In Shanghai Chinese the two registers are partly predictable from the preceding consonant, at least in initial position. We find a High register following aspirated and unaspirated voiceless stops, and a depressor register following voiceless depressor stops. Within the registers, we find three tonal melodies – level High, Rising and Falling – in long and short checked (e.g., with glottal coda) syllables. The registers are contrastive for tone, as not all melodies can combine with both registers: the depressor register has only Rising melodies (long vs. short).

The tone melody contrasts are neutralized in tone sandhi domains due to a process of assimilation. Non-initial syllables/morphemes in the domain lose their lexical tone (and tone register) – that is, the tone they bear when pronounced in isolation – and the melody of the initial morpheme (i.e., syllable) is distributed over the first two syllables of the tone sandhi domain. For example, as shown in the data below, if the initial syllable/morpheme has a lexical rising (MH) tone melody, the first two syllables of the sandhi domain will be M-H and any remaining syllables are realized with the default Low tone, whatever their isolation tones might be:[11]

(3) Shanghai tone neutralization in compounds (adapted, Zee and Maddieson 1980: 45-46, 61)

(a) *monosyllabic morphemes*

Morpheme	base tone	Gloss
/t^hi/	/HL/	'sky'
/ti/	/LM/	'earth'
/wəŋ/	/LM/	'studies'
/tɛ/	/LM/	'terrace'

(b) *bisyllabic compounds*

/HL/	/LM/	[H L]
/t^hi/	/ti/	[t^hi di]
'sky'	'earth'	'universe'

[10]See work like Y. Chen (2008, submitted), M. Chen (2000), Duanmu (1997), Selkirk & Shen (1990), Yip (2002), Zee & Maddieson (1980) and references therein for more detailed discussion of the phonetics and phonology of the Shanghai Chinese tone system.

[11]See Chen (2008) for a phonetic study of the f0 realization of the default L tone which leads to a somewhat different interpretation of how the non-neutralized tone contour of the initial syllable is realized in the sandhi domain from earlier proposals such as Selkirk & Shen (1990) and Duanmu (1993, 1997), among others.

(c) *trisyllabic compounds*

/HL/	/LM/	/LM/	[H M L]
/t^{h}i/	/wəŋ/	/tɛ/	[t^{h}i wəŋ dɛ]
‘sky’	‘studies’	‘terrace’	‘observatory’

Because the underlying tone is neutralized for all morphemes except the initial one in these tone sandhi contexts, the correlation between the occurrence of a depressor consonant and tone register of a following vowel is neutralized in non-initial position. However, because the voiceless depressors are phonetically voiced in non-initial syllables, the underlying register contrasts of non-initial syllables are maintained. That is, one could say that the loss of tonal contrasts in non-initial position is partially compensated for in Shanghai Chinese by medial depressor consonant voicing, as the underlying low tonal register is predictable from voicing (and tone melody is predictable from syllable shape).

4.2 Sketch of the Tone System of Zulu (and Other Nguni Languages)

Let us now sketch the essentials of the tone system of Zulu (shared with other Nguni Bantu languages) relevant to understanding the interaction of depressor consonants with tone, highlighting differences with Shanghai Chinese.[12] In Zulu, we find only one contrastive tonal register: a depressor register following the depressor consonants vs. a default non-depressed register following non-depressor consonants. Like other Bantu languages (Kisseberth and Odden 2003), Zulu is a level tone language, contrasting two level tones: High vs. default Low. The depressor register can combine with both High and default Low tones.

As in Shanghai Chinese, underlying (morphemic) tone contrasts are partially neutralized due to a process of assimilation. High tones spread (or shift) from underlyingly High-toned syllables (underlined in the data) up to the antepenult. This is illustrated in (4a), below. Depressor consonants not only affect the phonetic realization of High tones in some tonal contexts (see Fig. 1, above), they also interact with the tonal phonology. In Zulu (and other Nguni languages, like Xhosa (Cassimjee 1998) or Phuthi (Donnelly 2009)), they trigger the phonological process of Depressor High Tone Shift (DHTS), illustrated in (4b). This data shows that when the antepenult – the usual target of High tone shift – has a depressor consonant in the onset, the High tone shifts to the penult, resulting in a falling tone on that syllable. The data in (4c) illustrate another

[12] See work like Cassimjee & Kisseberth (1998, 2001); Donnelly (2009); Downing (1990, 2009); Khumalo (1981, 1987); Rycroft (1980) and references therein for detailed discussion and analysis of aspects of the Nguni tone systems, including the depressor effects sketched here.

correlation between depressor consonants and *f0* pitch lowering, namely, that DHTS is blocked if the penult also begins with a depressor consonant:

(4) Zulu tone neutralization in depressor and non-depressor contexts; all stems are Low-toned (Downing elicitation notes)

non-depressors

(a)	b<u>á</u>-yá-liima	'they are farming'
	b<u>á</u>-yá-heesha	'they are cutting grass'
	b<u>á</u>-yá-khúphuuka	'they are departing; ascending'
	b<u>á</u>-yá-námátheela	'they are sticking to'
	b<u>á</u>-yá-hlákúnísaana	'they are making each other weed'

depressors

(b)	*DHTS to penult*	
	b<u>á</u>-yá-**d**ilíika	'they are falling down' (*b<u>á</u>-yá-**d**íliika, without DHTS)
	b<u>á</u>-yá-**v**aléela	'they are closing someone out'
	b<u>á</u>-yá-**gij**imíisa	'they are running together'
	b<u>á</u>-yá-lá**nd**eláana	'they are following each other'

(c)	*DHTS blocked*		
	b<u>á</u>-yá-**bh**á**d**aala	'they are paying'	(*b<u>á</u>-yá-**bh**a**d**áala with DHTS)
	b<u>á</u>-yá-**gíj**iima	'they are running'	

As we can see from this data, in Nguni languages like Zulu – in contrast to Shanghai Chinese – the effect of depressors on tone register is not neutralized as the result of tone assimilation (i.e., tone spread or tone shift). On the contrary, active tonological processes like DHTS make the effect of depressor consonants on the register of following vowels particularly salient, whether the High tone is on that syllable underlyingly or due to assimilation. This confirms the phonetic findings shown in Fig. 1, above, namely, that the *f0* difference between vowels following the depressor and non-depressor register stops is consistently maintained (except in the Low-Low tone context). Unlike in Shanghai where laryngeal contrasts are always maintained by either tone register or consonant voicing, in Zulu we find no phonetic voicing of the depressor consonant, even in the Low-Low context where the laryngeal contrast between the three voiceless stop types is threatened because it is not signaled by *f0* cues.[13] That is, in Zulu, unlike Shanghai, voicing does not 'compensate' for loss of tone register realization.

---

[13] A possible explanation for this comes from the fact that register is not contrastive in Zulu. Further, the voiceless unaspirated vs. voiceless depressor contrast is marginal, due to the restricted number of morphemes beginning with voiceless unaspirated stops (Doke 1961: 8-9).

4.3 Phonological Differences Control Phonetic Implementation

How do these differences in the tonal systems help us explain the differences in the phonetic implementation of [slack voice] noted at the end of Section 3? First, we saw that there is a difference in the temporal domain of the depressor effect. In Shanghai Chinese, the depressor effect in medial position wanes earlier in the syllable, while in Zulu, the depressor effect lasts longer in the syllable in medial position. (This can be clearly seen by comparing Fig. 1 and 2, above.) We suggest that this difference has a straightforward phonological account, as tone has different domains of tonal realization in the two languages. In Shanghai Chinese, morphemes/syllables are lexically specified for tone, and the domain of tone realization is generally what work like Duanmu (1997), Chen (2000) and Yip (2002) argues is a two-syllable 'Foot'. In Zulu, the domain of tone contrast and tone realization is strikingly larger. Morphemes (often multisyllabic) are lexically specified for tone, and the domain of tone realization is the word. The different temporal domain of the depressors' effect on tone reflects the different temporal domain of tone in general.

There is also a difference in the realization of an *f0* – voicing trade-off for the depressor consonants. In both languages we superficially find a trade off. In Shanghai Chinese, depressors are phonetically voiceless in word-initial position where they correlate systematically with low tone register. In medial position, where they are voiced, they only variably lower *f0*. That is, in medial position, voicing could be said to compensate for a less consistent lowering effect on *f0*. In Zulu, depressor consonants are never voiced and they systematically affect *f0* regardless of their position in the word. That is, we find the opposite trade off: *f0* lowering could be said to compensate for systematic lack of voicing. The phonological account we propose for this difference is that in Shanghai Chinese, depressor consonants introduce a contrastive, lexically specified tonal register; tone sandhi neutralizes contrastive, lexically specified tones. Voicing the depressors partially maintains register and tonal contrasts which are lost in the medial (tone sandhi) context. That is, we find a trade-off in Shanghai Chinese between voicing and tone – albeit, just the opposite one suggested by Jessen and Roux (2002) – because the depressor register is contrastive for tone and would otherwise be neutralized in tone assimilation contexts. Phonetic voicing compensates for a phonological tone/register neutralization which also threatens to neutralize the three-way voiceless stop contrast. In contrast, depressor consonants in Zulu do not introduce a morphologically contrastive tonal register, and tone spread/shift does not target lexically specified tones. We therefore do not find the same trade-off in Zulu, because the depressor register is not contrastive for tone, and tone assimilation does not compromise either tonal contrasts or the depressor-non-depressor consonantal contrast. Instead, the *f0* difference which correlates with the depressor consonant consistently maintains the contrast among voiceless stops (except in the Low-Low context). Indeed, since voicing is never a phonetic cue to depressors in Zulu, *f0* lowering cannot really be said to compensate for lack of phonetic voicing, even though it might be considered a phonetic cue to a phonological [slack voice] feature.

5 How Much Variability in the Realization of a Phonological Feature?

Our proposal that the same feature [slack voice] can be adopted to characterize the depressor consonants in Shanghai Chinese and in Zulu, in spite of the different phonetic implementation of this feature in the two languages, naturally raises the question of how much phonologically motivated variability we can find in the cross-linguistic realization of a feature. As standard works like Keating (1988) and Kenstowicz (1994) note, phonological features must be abstract enough to define linguistically relevant categories ignoring some variation in phonetic implementation. Allowing for variability in phonetic implementation allows us to make generalizations that would be missed if a new feature were 'invented' to formalize every nuance of phonetic variability. In the spirit of Kingston and Diehl (1994), we have proposed that the single feature, [slack voice], can be adopted to characterize segments with different phonetic implementations, as phonetic implementation is controlled by the phonology. Indeed, we propose that possible variation in phonetic implementation is limited by requiring the phonetic differences to follow from differences in the phonological systems. Because the phonetic variation in the realization of depressor pitch lowering has a phonological motivation in Shanghai Chinese and Nguni languages, we prefer not to adopt Strazny's (2003) suggestion that a distinct feature [slack vocal cords] be used only for Nguni depressors. This would tie phonological features too closely to phonetic implementation, making cross-linguistic generalizations and comparisons difficult to formalize.

At the same time, as Keating (1988) and Kenstowicz (1994) also note, theories of phonology have standardly assumed that features are grounded in phonetics. Too much variability in the phonetic implementation of a phonological feature can bleach its definition of systematic, testable correlates. For this reason, we would like to distinguish [slack voice] from [voice]. While both features can have *f0* lowering as a phonetic correlate, we suggest that a distinct feature is required for depressor consonants like those found in Shanghai Chinese and Nguni Bantu languages that are voiceless and yet have *f0* lowering as a primary phonetic cue. As noted in Section 3, above, it is a mistake to automatically correlate *f0* lowering only with [voice], as [voice] is not the only (laryngeal) feature which correlates with pitch lowering. As a result, it is not plausible to characterize some consonants with the feature [voice] simply because they lower *f0*, especially when they are phonetically voiceless in pitch lowering contexts. Other laryngeal gestures may be involved, and labeling depressors with a distinct feature like [slack voice] should be a prod to further investigation of what these might be.

6 Conclusion

In sum, we have compared in detail the patterns of *f0* lowering and phonetic voicing of the voiceless depressors in Shanghai Chinese and Zulu. We show that although depressors in both languages may share the same phonological [slack voice] feature, as proposed by Jessen and Roux (2002), the phonetic implementation of this feature is clearly different in the two languages. We argue that these differences have a phonological explanation. In Shanghai Chinese, tones have a local (roughly, bisyllabic) domain of phonological interaction, while in Zulu they have a multisyllabic domain. In Shanghai Chinese tone assimilation potentially neutralizes both tone and consonant contrasts, while in Zulu it does not. We propose that it is these phonological differences that lead to the different interaction of tone and consonants in the two languages. Testing this proposal is obviously a topic for further research.

Appendix 1 – Zulu data set analyzed

All of the Zulu words in this list are verbs. The stem-initial syllables ('=' precedes the stem) contrast for the stop type: orthographic *ph, th, kh* are voiceless aspirated; *p, k* are voiceless unaspirated; *bh, d, g* are voiceless depressors; *b* is implosive. They also contrast for tone (High vs. default Low). In the recordings, these stems are preceded either by a sequence of High toned prefixes, *[bá-yá=]* 'they are X', or Low-toned prefixes, *[si-ya=]* 'we are X', to form complete one-word utterances. The data set is labeled to match the pitch track labels in Fig. 1 and 3.

Tone context and stop type	Zulu verb	Gloss
H_Aspirated_H	bá-yá = tháánda	'they like'
	bá-yá = phááthа	'they are carrying (in the hand)'
	bá-yá = phááka	'they are serving'
	bá-yá = kháála	'they are crying'
	bá-yá = thááka	'they are mixing medicines'
H_Unaspirated_H	bá-yá = pááka	'they are parking'
	bá-yá = péénda	'they are painting'
	bá-yá = póóka	'they are haunting'
H_Depressor_H	bá-yá = bhááia	'they are writing'
	bá-yá = bháánda	'they are plastering a hut with mud'
	bá-yá = dááya	'they are dying (cloth)'
	bá-yá = dáánsa	'they are dancing'
H_Aspirated_L	bá-yá = phaanda	'they are digging'
	bá-yá = khaaba	'they are kicking'
	bá-yá = phaahla	'they are daubing (mud)'

Tone context and stop type	Zulu verb	Gloss
H_Unaspirated_L	bá-yá = kaakwa	‘they are being surrounded’
H_Depressor_L	bá-yá = gaaya	‘they are grinding’
	bá-yá = bheeka	‘they are watching’
H_Implosive_H	bá-yá = báánda	‘they are cold’
	bá-yá = báába	‘they are hot-tempered’
	bá-yá = bíízwa	‘they are being called’
H_Implosive_L	bá-yá = baala	‘they are counting’
	bá-yá = baamba	‘they are catching; holding’

Tone sequence	Zulu verb (‘ = ’ precedes the stem)	Gloss
L_Aspirated_H	si-ya = tháánda	‘we like’
	si-ya = pháátha	‘we are carrying (in the hand)’
	si-ya = phááka	‘we are serving’
	si-ya = khááala	‘we are crying’
	si-ya = thááka	‘we are mixing medicines’
L_Unaspirated_H	si-ya = pááka	‘we are parking’
	si-ya = péénda	‘we are painting’
	si-ya = póóka	‘we are haunting’
L_Depressor_H	si-ya = bhááala	‘we are writing’
	si-ya = bhááanda	‘we are plastering a hut with mud’
	si-ya = dááya	‘we are dying (cloth)’
	si-ya = dáánsa	‘we are dancing’
L_Aspirated_L	si-ya = phaanda	‘we are digging’
	si-ya = khaaba	‘we are kicking’
	si-ya = phaahla	‘we are daubing (mud)’
L_Unaspirated_L	si-ya = kaakwa	‘we are being surrounded’
L_Depressor_L	si-ya = gaaya	‘we are grinding’
	si-ya = bheeka	‘we are watching’
L_Implosive_H	si-ya = báánda	‘we are cold’
	si-ya = báába	‘we are hot-tempered’
	si-ya = bíízwa	‘we are being called’
L_Implosive_L	si-ya = baala	‘we are counting’
	si-ya = baamba	‘we are catching; holding’

Appendix 2 – The Zulu consonant inventory (Schadeberg 2009)

		LAB	ALV	LAT	PAL	VEL	
1	aspirated voiceless plosives	ph	th			kh	
2	ejective plosives	p	t			k	
	prenasalized	mp	nt			nk	
3	ejective voiceless affricates		ts		tsh	kl	
	prenasalized				ntsh	nkl	
4	lax non-voiced plosives	bh̤	d̤			g̤	
	prenasalized voiced	mb̤	nd̤			ng̤	
5	lax non-voiced affricates				j̤		
	prenasalized voiced				nj̤		
6	implosive ɓ	b					
7	partially voiced ("soft") k					<k̬>	
8	voiceless fricatives	f	s	hl	sh		h
	prenasalized affricates	mf	ns	nhl			
9	voiced fricatives	v̤	z̤	dl̤			hh̤
	prenasalized affricates	mv̤	nz̤	ndl̤			
10	glides	w			y		
	depressor glides	<w>			<y̤>		
11	trilled and lateral sonorants		r	l			
12	nasal sonorants	m	n		ny	[ŋ]	
	depressor nasals	<m̤>	<n̤>				
13	voiceless clicks		c	x	q		
	prenasalized = nasal		n(k)c	n(k)x	n(k)q		
14	aspirated clicks		ch	xh	qh		
15	lax non-voiced clicks		gc̤	gx̤	gq̤		
	prenasalized voiced		ngc̤	ngx̤	ngq̤		

The spelling in this table follows modern orthography. It has the familiar arrangement where columns roughly correspond to places of articulation and rows to modes or manners of articulation. Some rows are further subdivided to show corresponding prenasalized consonants.

The shaded cells contain the depressor consonants, which are here also marked by a 'combining diaeresis below' (Unicode 1586). This marking is not part of standard orthography.

References

Bradshaw, Mary. 1999. *A Cross-linguistic Study of Consonant-Tone Interaction*. Ph.D. dissertation, The Ohio State University.

Brunelle, Marc. 2008. Speaker control in the phonetic implementation of Cham registers. Paper presented at TIE3, University of Lisbon, 16 September 2008.

Buell, Leston. 2005. *Issues in Zulu Verbal Morphosyntax*. Ph.D. dissertation, UCLA.

Cao, J., and Ian Maddieson. 1992. An exploration of phonation types in Wu dialects of Chinese. *Journal of Phonetics* 20: 77–92.

Cassimjee, Farida. 1998. *Isixhosa Tonology: An Optimal Domains Theory Analysis*. Munich: Lincom Europa.

Cassimjee, Farida, and Charles W. Kisseberth. 1998. Optimal Domains Theory and Bantu tonology: a case study from Isixhosa and Shingazidja. In Larry M. Hyman and Charles W. Kisseberth (eds.) *Theoretical aspects of Bantu Tone*, 33–132. Stanford, CA: CSLI.

Cassimjee, Farida, and Charles W. Kisseberth. 2001. Zulu tonology and its relationship to other Nguni languages. In Shigeki Kaji (ed.) *Proceedings of the Symposium, Cross-linguistic Studies of Tonal Phenomena: Tonogenesis, Japanese Accentology and Other Topics*, 327–359. Tokyo: ILCAA.

Chen, Matthew. 2000. *Tone Sandhi: Patterns across Chinese Dialects*. Cambridge: CUP.

Chen, Yiya. 2007. The phonetics and phonology of consonant-F0 interaction in Shanghai Chinese. Talk presented at the workshop 'Where Do Features Come From? Phonological Primitives in the Brain, the Mouth, and the Ear'. Paris, October 5, 2007.

Chen, Yiya. 2008. Revisiting the phonetics and phonology of Shanghai Tone Sandhi. Proceedings of Speech Prosody 2008. Campinas, Brazil.

Chen, Yiya. (submitted). How does phonology guide segment-*f0* interaction?

Chen, Yiya, and Carlos Gussenhoven (2008). Emphasis and tonal implementation. *Journal of Phonetics* 36: 724–746.

Clements, G.N. 2003. Feature economy in sound systems. *Phonology* 20: 287–333.

Clements, G.N., and Rajesh Khatiwada. 2007. Phonetic realization of contrastively aspirated affricates in Nepali. *Proceedings of ICPhS XVI* (Saarbrücken, 6–10 August 2007), 629–632.

Doke, Clement M. 1961. *Textbook of Zulu grammar*. London: Longmans, Green and Co.

Donnelly, Simon. 2009. Tone and depression in Phuthi. In Michael Kenstowicz. (ed.) *Data and Theory: Papers in Phonology in Celebration of Charles W. Kisseberth*, Special issue, *Language Sciences* 31: 161–178.

Downing, Laura J. 1990. Local and metrical tone shift in Nguni. *SAL*21: 261–317.

Downing, Laura J. 2009. On pitch lowering not linked to voicing: Nguni and Shona group depressors. With an Appendix by Thilo C. Schadeberg. In Michael Kenstowicz. (ed.) *Data and Theory: Papers in Phonology in Celebration of Charles W. Kisseberth*, Special issue, *Language Sciences* 31: 179–198.

Downing, Laura J., and Bryan Gick. 2005. Voiceless tone depressors in Nambya and Botswana Kalang'a. *BLS* 27 (2001): 65–80.

Duanmu, San. 1993. Rime Length, Stress, and Associated Domains. *Journal of East Asian Linguistics* 2: 1–44.

Duanmu, San. 1997. Recursive constraint evaluation in Optimality Theory: evidence from cyclic compounds in Shanghai. *NLLT* 15: 465–508.

Gandour, J., Potisuk, S., and S. Dechongkit. 1994. Tonal coarticulation in Thai. *Journal of Phonetics* 22: 477–492

Giannini, Antonella, Massimo Pettorino, and Maddalena Toscano. 1988. Some remarks on Zulu stops. *AAP* 13: 95–116.

Hombert, Jean-Marie. 1978. Consonant type, vowel quality, and tone. In Victoria A. Fromkin (ed.) *Tone: A Linguistic Survey*, 77–111. New York: Academic Press.

Jessen, Michael, and Justus C. Roux. 2002. Voice quality differences associated with stops and clicks in Xhosa. *Journal of Phonetics* 30: 1–52.

Jun, Sun-Ah. 1996. Influence of microprosody on macrosprosody: a case of phrase initial strengthening. *UCLA Working Papers in Phonetics* 92: 97–116.

Keating, Patricia A. 1988. The phonology-phonetics interface. In Frederik J. Newmeyer (ed.) *Linguistics: The Cambridge Survey*, vol. I. *Linguistic Theory: Foundations*, 281–302. Cambridge: Cambridge University Press.

Kenstowicz, Michael. 1994. *Phonology in Generative Grammar*. Cambridge, Mass.: Blackwell.

Khumalo, J.S.M. 1981. Zulu tonology, Part 1. *African Studies* 40: 53–130.

Khumalo, J.S.M. 1987. *An Autosegmental Account of Zulu Phonology*. PhD dissertation, University of the Witwatersrand.

Kingston, J., and R. L. Diehl. 1994. Phonetic knowledge. *Language* 70: 419–454.

Kisseberth, Charles, and David Odden. 2003. Tone. In Derek Nurse and Gérard Philippson (eds.) *The Bantu Languages*, 59–70. London: Routledge.

Ladefoged, Peter, and Ian Maddieson. 1996. *The Sounds of the World's Languages*. Oxford: Blackwell.

Lee, Seunghun. 2008. *Consonant-Tone Interaction in Optimality Theory*. Ph.D. dissertation, Rutgers University.

Maddieson, Ian. 2003. The sounds of the Bantu languages. In Derek Nurse and Gérard Philippson (eds.) *The Bantu Languages*, 15–41. London: Routledge.

Matisoff, James A. 1973. Tonogenesis in Southeast Asia. In Larry M. Hyman (ed.) *Consonant types and tone*, *SCOPIL* 1: 71–95.

Ren, N. 1992. *Phonation types and consonant distinctions: Shanghai Chinese*. PhD dissertation, The University of Connecticut.

Rycroft, D.K. 1980. *The 'Depression' Feature in Nguni Languages and Its Interaction With Tone*. Communication no. 8, Dept. of African Languages, Rhodes University, Grahamstown, R.S.A.

Schadeberg, Thilo C. 2003. Derivation. In Derek Nurse and Gérard Philippson (eds.) *The Bantu Languages*, 71–89. London: Routledge.

Schadeberg, Thilo C. 2009. Appendix: On the origin of Zulu depressor consonants. In Michael Kenstowicz (ed.) *Data and Theory: Papers in Phonology in Celebration of Charles W. Kisseberth*, Special issue, *Language Sciences* 31: 192–197.

Selkirk, Elisabeth O., and Tong Shen. 1990. Prosodic domains in Shanghai Chinese. In Sharon Inkelas and Draga Zec (eds.) *The Phonology-Syntax Connection*, 313–337. Chicago: CSLI.

Silva, D. 2006. Acoustic evdience from the emergence of tonal contrast in contemporary Korean. *Phonology*23: 287–308.

Strazny, Philipp. 2003. Depression in Zulu: tonal effects of segmental features. In Jeroen van de Weijer, Vincent J. van Heuven and Harry van der Hulst (eds.) *The Phonological Spectrum, vol 1: Segmental Structure*, 223–239. Amsterdam: John Benjamins.

Svantesson, Jan-Olof, and David House. 2006. Tone production, tone perception and Kammu tonogenesis. *Phonology* 23: 309–333.

Tang, Katrina E. 2008. *The Phonology and Phonetics of Consonant-Tone Interaction*. Ph.D. dissertation, UCLA.

Traill, A. 1990. Depression without depressors. *South African Journal of African Languages*10: 166–172.

Traill, A., J.S.M. Khumalo, and P. Fridjhon. 1987. Depressing facts about Zulu. *African Studies* 46: 255–274.

Wright, Richard. 1992. The effect of implosives on fundamental frequency in SiSwati. Paper presented at ACAL 23, 26–29 March 1992.

Xu, G., and Z. Tang (eds.). 1988. *Shanghai Fangyan Gaikuang*. Shanghai: Shanghai Educational Press.

Xu, Y. 1997. Contextual tonal variations in Mandarin. *Journal of Phonetics* 25: 61–83.

Yip, Moira. 2002. *Tone*. Cambridge: Cambridge University Press.

Zee, Eric, and Ian Maddieson. 1980. Tones and tone sandhi in Shanghai: Phonetic evidence and phonological analysis. *Glossa* 14: 45–88

Tonal and Non-Tonal Intonation in Shekgalagari

Larry M. Hyman and Kemmonye C. Monaka

1 Introduction

It is common knowledge that tone languages can have F0-based intonation, including H and L intonational tones (or 'intonemes'): 'Even in languages with elaborate omnisyllabic tone systems, intonation certainly exists as a phenomenon independent of tone.' (Matisoff 1994: 116) 'Most tone languages will have some form of structural intonation.' (Gussenhoven 2004: 45) The occurrence of intonational tones alongside lexical ones is, however, not without potential complications. Word-level tones show three degrees of hospitality (or hostility) towards F0 intrusions at the phrase- or utterance level:

(i) *Accommodation* ('peaceful coexistence'), whereby the terrain is divided up somehow such that the lexical and intonational tones minimally interact. One instantiation of this occurs in certain Otopamean languages of Mexico, which restrict their lexical tone contrasts to pre-final syllables, reserving word-final syllables for intonational contrasts. An example is Mazahua, which contrasts /H/, /L/ and /HL/ tones: 'The pitches of all syllables which do not immediately precede word space are those of the tonemic system. The pitch of any syllable immediately preceding word space is part of the intonemic system' (Pike 1951: 101). The 'intonemes' which are distinguished in Mazahua are identified in Table 1 in (1).

It should be noted that while intonemes can be combined, they never go beyond being distributed on one syllable. This is not to say that there is no interaction. Thus, while Mazahua lexical and intonation tones generally accommodate each other by staying on their respective syllables, Pike (1951: 103) further explains:

L.M. Hyman (✉)
Department of Linguistics, University of California, CA 94720-2650, Berkeley, USA
e-mail: hyman@berkeley.edu

S. Frota et al. (eds.), *Prosodic Categories: Production, Perception and Comprehension*, Studies in Natural Language and Linguistic Theory 82,
DOI 10.1007/978-94-007-0137-3_12,

Table 1 Intonemes in Mazahua

(1)	Intoneme	Meaning	Intoneme	Meaning
	L%	‘colorless finality’	MH%	‘surprise’
	H%	‘is that what you said/mean?’	ML%	‘anger, disgust’
	M%	‘something is expected to follow’	H:L%	‘calling, shouting’

> A simple stem is made up of two syllables, one of which is a root and the other a stem formative. The root contains the toneme, and the stem formative normally carries an intoneme. When a compound is formed of stems whose stem formatives are composed of ʔV, the stem formatives are dropped and one of the roots now occurs word final. In word-final position, the toneme of the root becomes obliterated and an intoneme takes its place. In this way the intonation-character of word-final syllables has overpowered the earlier tonemic character of the old penultimate syllable.

The mutual accommodation of lexical and intonational tones is thus imperfect in Mazahua. While an intoneme can ‘overpower’ a lexical tone, the other logical outcome is where a lexical tone blocks intonation. These constitute the two remaining types of interaction between lexical tones and intonemes:

(ii) *Submission* (‘surrender’), whereby the intonational tones invade and override the lexical tones. A rather striking case of this occurs in Coreguaje (Tukanoan; Colombia), where in isolation CVCV noun tones merge as L-HL with statement intonation and as H-L with question intonation: ‘...we found that in certain frames there were four contrasting sets, but in isolation phrase stress completely neutralized the contrasts, at least in CVCV nouns’ (Gralow 1985: 3). As seen in Table 2 in (2), CVV noun tones also merge except for /LL/ nouns, which remain distinct under statement intonation:

It would appear that the statement and question intonemes are LHL% and HL%, respectively, although more information would be needed to confirm this.

(iii) *Avoidance* (‘blockade’) constitutes the third type of interaction between lexical tones and intonation. In this case intonation is minimized, perhaps limited to Ladd’s (1980, 1996) ‘paralinguistic’ modulations (pitch range and pitch interval adjustments, etc.) One possibility is incomplete avoidance whereby one or more lexical tones override one or more intonemes. A second is complete avoidance, where the tone system does not tolerate any intonemes. If intonational tones cannot be exploited, their common functions may be fulfilled by something else, e.g. by particles: ‘... omnisyllabic tone languages

Table 2 Tone Patterns in Corejuage

(2)	CVCV:	Basic form	statement	question	CVV:	Basic form	statement	question
		H-H	L-HL	H-L		HH	HL	HL
		H-L	L-HL	H-L		HL	HL	HL
		L-L	L-HL	H-L		LL	LH	HL
		L-H	L-HL	H-L				

typically have a repertoire of particles whose only job is to convey the emotion or affect of the speaker—syllabic exclamation points, as it were' (Matisoff 1994: 118).

With the above potential tone-intonation relations established, we now come to the following two questions:

(3) a. Can a language do without structural intonation?
 b. Can an utterance lack intonation?

Concerning the first, many researchers have assumed that intonation is a universal:

> Every human language has both an intonational system and a nonintonational system.... (Hockett 1963: 19)
>
> Intonation is universal first of all because every language possesses intonation.... Intonation is universal also because many of the linguistic and paralinguistic functions of intonation systems seem to be shared by languages of widely different origins. (Hirst and Di Cristo 1998: 1) (but see Ladd 1996, Ch. 4)
>
> Languages described as having 'no intonation'... or 'no contrastive pitch patterns'... are still admitted to have changes in pitch corresponding to the fluctuations of emotion. (Bolinger 1978: 475)

However, the question is not whether all languages have utterance-level F0 modulations such as raising vs. lowering of pitch level or expanding vs. compressing of pitch intervals, but rather whether there are languages which lack 'structural' intonation, i.e. categorical intonational pitch features or intonemes. The strongest limiting cases are probably 'monosyllabic languages' with highly developed tone systems such as the five levels of Dan (Mande; Ivory Coast) (Bearth and Zemp 1967, Vydrine and Kességbeu 2008:10) or Wobe (Kru; Liberia) (Bearth and Link 1980, Singler 1984), which contrasts fourteen tones (four levels, ten contours) on monosyllables. In such languages intonational tones would not only have to cope with the strong competition from so many lexical tones, but also with analytic ambiguities. If, for example, such a highly tonal language marked questions by a final high pitch, this could be evidence of a H% 'intoneme' phonologizing Gussenhoven's (2004) frequency code. However, it could also be an accidental 'tonal morpheme' derived from an old monosyllabic interrogative particle that has lost its segments, but whose *H is preserved. The same ambiguity would be present if a question or any other utterance type were marked instead by a final low pitch.

If we assume for the purpose of discussion that intonation is universal, the question in (3b) then asks whether specific utterances in an individual language can be intonation-less. In other words, if a language has intonation, does this mean that all utterances are marked by an intonation, whether structural or 'paralinguistic'? Again, this possibility would seem most likely to arise in complex tone languages. If absence of intonation does occur, would intonation-less utterances represent a kind of default, or could the absence of intonation itself signal specific pragmatic functions? As we shall see, this very question arises in Shekgalagari.

As mentioned above, languages with ‘omnisyllabic tone’ (Matisoff 1994) may choose to use particles and perhaps avoid structural intonation altogether. While it has yet to be established that tone languages do in fact make greater use of particles than non-tonal languages, they need not give up on intonation altogether. There exists an alternative intonational strategy: use features other than tone. This is exactly the situation in Shekgalagari. In the following sections we shall first establish the basic tone system of the language, followed by a systematic description of the intonational marking of different utterance types in Shekgalagari. We shall see that the marking of intonation goes well beyond structural tones or pitch adjustments, thereby raising questions concerning the inventory of intonational features and the nature of intonation itself.

2 Shekgalagari Basic Tonology

Shekgalagari is a Bantu language of the Sotho-Tswana group designated as S.30 by Guthrie (1967–1971). Although sometimes lumped with Setswana, it is a separate language (Janson 1995), spoken by an estimated 272,000 speakers in Botswana (RETENG 2006) as well as a smaller number of speakers in Namibia. Previous research on the language has been relatively limited but includes Krüger and du Plessis (1977), Dickens (1984, 1986a, b), Neumann (1999) and Monaka (2005a, b). Recent research includes a grammar (Lukusa and Monaka 2008), a lexicon (Monaka, in preparation), and a detailed description of the tone system (Crane 2008, 2009a, b). The material presented in this paper are based on the speech of the second author, which differs only slightly (and in irrelevant details) from previous documentation of the language.

The basic properties of the tone system are as follows: The underlying system is characterized by a binary contrast, probably best analyzed as /H, Ø/ rather than /H, L/ (Crane 2008, 2009a, b). The surface system consists of four pitch levels: H, L, ꜜH and LꜜL, all but the last of which are level. While ‘non-automatic’ downstepped ꜜH is contrastive after another (ꜜ)H tone, there is no perceptible ‘automatic downstep’ or ‘downdrift’ in $H\text{-}L^n\text{-}H$ sequences. Unlike many other tone systems, L tones are level in pitch, even before pause, thereby sometimes giving the impression of a mid tone level. Two contour tones, HL and LꜜL, occur only on a lengthened penultimate syllable. LꜜL consists of a L tone falling to an even lower pitch level. Finally, there are no rising tones in the language. While there are occasional long vowels which have a L to H rise, there is independent evidence that these should be analyzed as heterosyllabic sequences of identical vowels ($V_i.V_i$). In citing examples, L tone is unmarked, H tone is marked by an acute accent (á), HL falling tone by a circumflex (â:), and the LꜜL falling tone by a grave accent (à:).

We begin in (4) by presenting the four logical combinations of H and L tones on bisyllabic noun stems, as they are pronounced utterance-medially:

(4)	prefixless		prefixed	
L-L	nama	'meat'	mʊ-lɪmi	'farmer'
L-H	nawá	'bean'	ma-rumé	'greetings'
H-L	lóri	'lorry'	mʊ-nʊ́na	'man'
H-H	nárí	'buffalo'	mu-rérí	'preacher'

As seen, nouns may be prefixed or not, the L tone prefixes marking the typical Bantu noun classes. The reason for citing utterance-medial outputs will become apparent, but is done basically to avoid the tonal complications which accompany utterance-penultimate vowel lengthening (see (8)).

As seen in (5), in the infinitive, which is marked by a *xʊ-* prefix, verb stems exhibit three different tone patterns independent of the length of the verb stem:

(5) all L (1291)		two Hs (1484)		one H (590)	
k-a	'mention'			gy-á	'eat'
bal-a	'count'	bɔ́n-á	'see'	lɔ́r-a	'dream'
lɛlɛk-a	'chase'	rʊ́mɛ́l-a	'send'	lález-a	'invite'
xalaleʒ-a	'praise'	sérímuʃ-a	'reveal'	bʊ́tsʊlʊs-a	'avenge'
makyʊrʊlʊl-a	'unstick'	xáqʊ́lʊʃɛlw-a	'remember'	bʊ́tsʊlʊseʒ-a	'pay back'

As seen, monosyllabic stems can be L or H, while bisyllabic and longer stems show three patterns: all L, a H on the first two syllables, and a H on the first syllable followed by all low. The numbers in the headings indicate how many lexical entries were found of each tone pattern in Monaka (in preparation), not counting the 11 monosyllabic H verbs which occur in the language. As seen, the one H pattern is distinctly in the minority. This is because there is a general rule of bounded H tone spreading (HTS) which has the effects shown in (6) (Crane 2008):

(6)	bɔn-a	rʊmɛl-a	serimuʃ-a	xaqʊlʊʃɛlwa
	H	H	H	H

As originally pointed out by Dickens (1984), the one-H forms are exceptions to HTS deriving from loss of the Proto-Bantu (PB) vowel length (*VV):

(7) Proto-Bantu	Shekgalagari		Proto-Bantu	Shekgalagari	
*bón-a	bɔ́n-á	'see'	*dóot-a	lɔ́r-a	'dream'
*dʊ́m-a	lʊ́m-á	'bite'	*dáad-a	lál-a	'sleep'
*tʊ́k-a	rʊ́x-á	'insult'	*bʊ́ʊd-i-a	bʊ́ʒ-a	'ask'

The first two columns show that PB H tone *CVC-a stems surface as H-H, while the forms to the right show that PB *CVVC-a stems are realized as H-L. What this means is that pre-Shekgalagari (**lɔ́ɔr-a*) 'dream' first becomes (*lɔ́ɔ́r-a*) by HTS, and then lɔ́r-a by vowel-shortening. The same H-L pattern is found on native noun stems, but also in borrowings, e.g. *lóri* 'lorry'.

As mentioned, forms have thus far been cited as they appear in medial position, i.e. when not immediately preceding pause. The reason for this is that a pause-penultimate vowel is lengthened in declarative utterances, including citation forms. Thus, compare the pre-pausal realizations of the nouns in (8) with the corresponding medial forms in (4).

(8)	input			prefixless		prefixed	
	L-L	→	L↓L:-L	nà:ma	'meat'	mʊ-lì:mi	'farmer'
	L-H			na:wá	'bean'	ma-ru:mé	'greetings'
	H-L			ló:ri	'lorry'	mʊ-nʊ́:na	'man'
	H-H	→	HL:-L	nâ:rɪ	'buffalo'	mu-rê:ri	'preacher'

As indicated, and as summarized in (9), the penultimate lengthening can have an effect on the tones of the last two syllables before pause:

(9) a. no effect other than lengthening if the last two syllables differ in tone

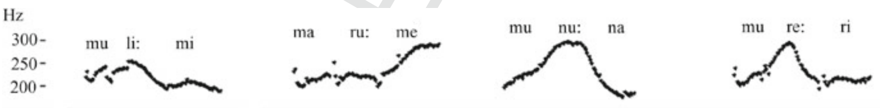
L-H	→	L:-H	:	ma-rumé	→	ma-ru:mé 'greetings'
H-L	→	H:-L	:	mʊ-nʊ́na	→	mʊ-nʊ́:na 'man'

b. pitch of the penult falls if the last two syllables have the same input tones

L-L	→	L↓L:-L	:	mʊ-lɪmi	→	mʊ-lɪ:mi 'farmer'
H-H	→	HL:-L	:	mʊ-rérí	→	mʊ-rê:ri 'preacher'

Representative pitch traces of the above four nouns are provided in Fig. 1 in (10), with thanks to Keith Johnson for his guidance:

(10) Fig. 1 Pitch Tracks and Hz Values by Syllable for the Four Nouns in (9)



mʊ	lì:	mi	ma	ru:	me	mʊ	nʊ́:	na	mu	re:	ri
234	253-206	208	215	223	291	207	298	200	213	294-218	216

Before proposing an analysis of the tone changes in (9b), note that the alternations are also observed on the verb stems of the infinitive when they appear before pause. The forms in (11) should thus be compared with those in (5):

(11)	all L		two Hs	(cf. (5))	one H	
	(xʊ̀:-) k-a	'mention'			(xʊ:-) gy-á	'eat'
	bà:l-a	'count'	bɔ̂:n-a	'see'	lɔ́:r-a	'dream'
	lɛlɛ̀:k-a	'chase'	rʊ́mɛ́:l-a	'send'	lálè:ʒ-a	'invite'
	xalalè:ʒ-a	'praise'	sérímù:ʃ-a	'reveal'	bʊ́tsʊlʊ̀:s-a	'avenge'
	pakyʊlʊxà:n-a	'come apart'	xáqʊ́lʊʃɛ̂:lw-a	'remember'	bʊ́tsʊlʊsè:ʒ-a	'pay back'

As indicated, when the verb stem is monosyllabic, the infinitive prefix is lengthened to *xʊ:-*.

Our proposal to account for the observed tone changes is that a L% intonational tone links to the second mora of the lengthened penult when the final two syllables are Ø-Ø or H-H:

```
(12)  a.     L%                                 b.     L%
             |                                         |
           naama   [nà:ma]   'meat'                  naarɪ   [nâ:rɪ]   'buffalo'
                                                      \/
                                                      H
```

In (12), we assume that the lexical tonal contrast is between /H/ and /Ø/. The L% which is shown on a separate tier is the intonational tone which has an audible effect only when the last two tones are identical. Since L% represents a tone lower than /Ø/, the result will be a L to ↓L falling tone in (12a). A striking fact about the the penultimate L↓L contour is that Ls which precede it are realized higher than Ls which precede H. Thus, one can see in (10) that the prefix *mʊ-* is realized higher on *mʊ-lì:mi* 'farmer' than on *mʊ-nʊ́na* 'man'. In longer forms, a sequence of Ls is audibly raised to anticipate the L↓L fall. It is striking that H-H alternates with HL:-L. The derivation we propose is in (13).

(13) H-H → HL:-H → HL:-L (HL%-Ø)

First the L% splits up the two Hs of the H:-H sequence to produce HL:-H. Subsequently, the final H is delinked, thereby creating the flat L pitch on the last syllable. Evidence for the intermediate step is seen from Ikalanga, a nearby Bantu language of the Shona group, which has corresponding alternations such as *túmá* ~ *tû:má* 'send' (Hyman and Mathangwane 1998; Mathangwane 1999). Although we have no evidence that L% is present when the last two tones are H-L or L-H, note that we may allow L% to link to final syllable of prepausal H-L and to the penultimate syllable of prepausal L-H. If correct, the generalization would be that L% links to the penult unless the form ends H-L, in which case L %; links to the final syllable. As an alternative, we considered the following: If H-L were represented as /HL-Ø/, it could undergo the following multistep derivation prepausally: HL-Ø → H-L (→ HL:-L?) → H:-L. This would keep /HL-Ø/ distinct from /H-Ø/, which could become HL:-Ø directly. While it may seem autosegmentally unusual to derive HL:-H from intermediate H-H, the alternative analysis requires the counterintuitive assignment of intonational length + L% before applying word-bounded HTS. Instead, Ikalanga justifies the analysis in (13).

Having illustrated prepausal lengthening and its tonal consequences on citation forms, we are now ready to consider the full range of intonation in Shekgalagari. In the following discussion we shall refer to the penultimate lengthening + L% tone as PLL.

3 Shekgalagari Intonation Types

As in other studies of intonation, it is necessary to establish both the prosodic features marking different intonations as well as the utterance types in which they occur. As seen in Table 3 in (14), many Bantu languages have penultimate lengthening. In the left column we have indicated the utterance types or functions marked by penultimate lengthening in one or more of the above languages. Since grammars rarely go into such detail, the above table was made possible only by generous personal communications from our Bantuist colleagues, specifically Malillo Machobane and Katherine Demuth (Sesotho), Philippe Ngessimo Mutaka (Kinande), Joyce Mathangwane (Ikalanga), Galen Sibanda (Ndebele), and Sam Mchombo and Al Mtenje (Chichewa). As indicated, none of the cited languages restricts penultimate lengthening like Shekgalagari. The fact that PLL also is marked by L% is clearly duplicated in other languages of the Sotho-Tswana group, as well as by Ikalanga. In other languages such as Kinande and Chichewa penultimate lengthening is not accompanied by an intonational tone in ordinary declaratives.

Table 3 Penultimate lengthening in six Bantu languages

(14)	Shekgalagari	Sesotho	Ikalanga	Kinande	Ndebele	Chichewa
Declaratives	+	+	+	+	+	+
Yes-No Q	–	–	+	–	+	+
WH Q	–	–	+	–	+	+
Ideophones	–	–	–	–	+	+
Paused lists	–	+	–	+	+	+
Imperatives	–	+	+	+	+	+
Hortatives	–	+	+	+	+	+
Vocatives	–	±	+	+	+	+
Exclamatives	–	–	+	+	+	+
1σ word	–	+	+	+	+	+

We now briefly illustrate the presence vs. absence of PLL in each of the utterance types listed in (14). As seen in (15), PLL occurs before pause in declarative indicatives, including citation forms:

(15)	a.	ri-nâ:rɪ	'buffalos'	a-bal-a ri-nâ:rɪ	'he is counting buffalos'
	b.	xʊ-bɔ̂:n-a	'to see'	a-bɔ́n-á mʊ-lì:mi	'he sees the farmer'

Unlike the other Bantu languages characterized in (14), (15) represents the only utterance types in which PLL is required in Shekgalagari. Failure to lengthen would unambiguously result in these forms being interpreted as yes-no questions:

(16)	a.	ri-nárɪ́	'buffalos?'	a-bal-a ri-nárɪ́	'is he counting buffalos?'
	b.	xʊ-bɔ́n-á	'to see?'	a-bɔ́n-á mʊ-lɪmi	'does he see the farmer?'

Correspondingly, the examples in (17) show that there is no PLL in WH questions (the downstep in (17a) and elsewhere is irrelevant for our purposes—see Crane 2008, 2009a, b):

(17)	a.	ri-nárí zhé ↓ríhí	‘which buffalos?’	
	b.	a-bal-a ʃiŋ́	‘what has he just counted?’	(ʃi.ŋ́ = bisyllabic with L-H tone)
	c.	xʊ-bɔ́n-a ányí	‘to see who?’	
	d.	ányí a-bɔ́n-á mʊ-lɪmi	‘who has just seen the farmer?’	

While there are other Bantu languages which suspend penultimate lengthening in questions, Shekgalagari is thus far the only known to disallow PLL in imperatives (cf. Hyman 2009):

(18)	a.	bal-á	‘count!’
	b.	bal-á ↓rí-nárí	‘count the buffalos!’
	c.	bɔ́n-a	‘see, look!’
	d.	bɔ́n-á mʊ-lɪmi	‘see the farmer!’

The same is true of hortatives:

(19)	a.	á ↓hí-bál-ɛ	‘let’s count!’
	b.	á ↓hí-bál-ɛ ri-nárí	‘let’s count the buffalos!’
	c.	á ↓bá-bɔ́n-ɛ	‘let them see!’
	d.	á ↓bá-bɔ́n-ɛ mʊ-lɪmi	‘let them see the farmer!’
	e.	á mʊ-lɪmi a-w-ɛ	‘let the farmer fall!’

PLL is likewise not found in vocatives and terms of address:

(20)	a.	Mʊnaká	‘Monaka!’
	b.	ntó Gabalʊxʊ́ŋ	‘come here, Ghabalogong!’
	c.	tatɛ́	‘father!’
	d.	ɛɛ mmá	‘yes, ma’am’ (m.má = two syllables with L-H tone)

The data in (21) show that there is no PLL in exclamatives, which use the same *á* marker as hortatives:

(21)	a.	á ↓ʃɪ́-xʊ́lʊ́	‘what a situation!’
	b.	á ↓ʃɪ́-sʊ́lʊ́	‘what a bargain!’
	c.	á ↓ʃɪ́-tʃʊ́tʃʊ	‘what an idiot!’
	d.	á ↓ʃɪ́-tʃʊ́tʃʊ ʃá mʊ́-khyʊ	‘what an idiot of a person!’

The above constitutes the list of utterance types where the prepausal forms without PLL are identical to how they would appear in utterance-medial position. Two additional utterance-types also block PLL but add an intonational mark of their own. First, ideophones have a short penult. In addition, their pre-pausal vowel undergoes final devoicing (FD):

(22) a. y-á-rɪ bílu̥ 'it (fish) appeared suddenly out of water' (*it went BILU*)
b. a-rɪ bítsɪ̥ 'he left in a hurry' (*he went BITSI*)
c. l-á-rɪ phátsi̥ 'lightening flashed' (*it went PHATSI*)
d. a-rɪ tshíki̥ 'it's cold, I'm feeling cold' (*it went TSHIKI*)

As in many Bantu languages, there is a general verb, here *-rɪ* 'say', which is used with ideophones. The equivalent in English is to use the verb 'go', as indicated in the parenthetical paraphrases to the right of the above examples. What is important is that the final vowel must be devoiced in the declarative (see below for the corresponding interrogatives). We will argue below that ideophone devoicing is intonational.

Like ideophones, the internal members of 'paused lists' are not subject to PLL, but undergo final lengthening (FL):

(23) a. a-bal-a ri-nama: . . . ri-nawá: . . . lí ri-nâ:rɪ
'he's counting meats. . . beans. . . and buffalos'
b. a-bɔ́n-á lʊ-rʊli: . . . malílɪ: . . . lí mʊ-rɪ:ri
'he sees dust. . . rubbish. . . and hair'

For there to be such lengthening, it is obligatory that there be a pause after each of the listed items. In other languages, such paused lists are often marked by a final rising intonation with possible lengthening. This brings us to the following observation: Recalling that the declarative not only lengthens the penultimate vowel, but also assigns a L% tone to its second mora, it is striking that interrogatives, imperatives, hortatives, vocatives, exclamatives, ideophones, and paused lists are all suspended and/or vivid speech act types where speakers might be expected to raise their voice. Could this be related to the fact that they all resist the fall-creating L% tone—a blocking effect attributable to the Frequency Code (Gussenhoven (2004: 82)?

Before summing up this section, it is necessary to consider one last relevant environment: Shekgalagari differs from related languages in not assigning PLL when the prepause word is monosyllabic:

(24) a. ri-nárí ʒé 'these buffalos'
b. ʊ-bat-a ʃé 'he wants this one'
c. a-rí-bál-a ʒwá 'he has counted them in this way'
d. qa-rɪ ʊ́-ʃíkíl-a = xɔ́ thɛ́ 'I say, you really move around'
I-say you-go = infl really

In related Bantu languages, the final vowel of the preceding word would be lengthened. What this shows is that PLL is sensitive to word boundaries (cf. the Appendix).

To summarize, we have seen four different intonational patterns before pause:

(25)	a.	PLL	:	declaratives, citation forms
	b.	FD: final devoicing (no PLL)	:	ideophones
	c.	FL: final lengthening (no PLL)	:	paused lists
	d.	Ø (none of the above)	:	yes-no questions, WH-questions, imperatives, hortatives, vocatives, exclamatives, 1σ words

Finally, for clarity, it should be noted that the above intonations cannot be combined: It is totally ungrammatical for the last two syllables of any utterance to undergo PLL + FD, PLL + FL, or FD + FL. The above raises the following two questions:

(26) a. Are all Shekgalagari utterances marked by an intonation?
b. Which pattern is unmarked, the default: PLL or Ø?

The answer to the first question depends on how we interpret (25d): the absence of PLL, FD or FL. If Ø is an intonation which is actively assigned, then there are four intonations in Shekgalagari: PLL, FD, FL, Ø. If Ø is not an actively assigned intonation, then utterances not marked by PLL, FD or FL can in fact exist without an intonation. The second question which concerns questions of markedness is related to the first. The problem in Shekgalagari is that phonological and pragmatic markedness are at odds with each other: Declaratives and citation forms are PRAGMATICALLY unmarked speech acts, but are PHONOLOGICALLY marked by the intrusive mora and L% feature. On the other hand, a short penult is phonologically unmarked but pragmatically marked, thus assigned to questions, imperatives, vocatives etc. We would like to answer 'yes' to (26a) and assume that Ø is an intonation, but that PLL is pragmatically unmarked. To see how this might work out it is necessary to investigate how an utterance is realized which qualifies for more than one intonation. This is taken up in the next section.

4 Competing Intonations in Shekgalagari

The question we address in this section is: What happens when a construction qualifies for more than one intonation? Which one wins? For example, what happens if a question ends in an ideophone: Will the ideophone undergo FD, or will it be marked by Ø? If the latter, this gives further evidence that Ø is an actively assigned intonation. Two logical resolutions of such conflicts have occurred to us: (i) There could be a fixed hierarchy of utterance types and their intonations. (ii) There could be variation, with the outcome depending on the intention of the speaker or on the relative importance that the speaker gives to each of the inputs. To some extent both properties are found in Shekgalagari.

Let us first examine whether a fixed hierarchy is possible. A first approximation which will now be examined is presented in (27).

(27)	Yes-No, WH-Q >>	Ideo >>	Imper, Hort, Voc, Excl >>	List >>	Decl
	Ø	FD	Ø	FL	PLL

The first line in (27) hierarchizes the different utterance types, while the second line provides a reminder of the intonation associated with each type. While it is not logically possible to combine all utterance types (e.g. an ideophone cannot be used in a vocative utterance), vocative and exclamatory utterance types have been grouped together with imperatives and hortatives with which they seem otherwise to pattern. While we will next illustrate the implied conflict resolutions, the following should be noted concerning (27): (i) Ø needs to be split up in the hierarchy; (ii) interrogative Ø can override any other intonation; (iii) declarative PLL never overrides anything (but cf. emphatic PLL below).

Let us consider some of the conflict resolutions implied in (27). To begin, the utterances in (28) illustrate how interrogative Ø can suspend the final devoicing on ideophones:

(28) Yes-No, Wh-Q >> Ideo (Ø >> FD)

a. y-á-rɪ bílʊ? 'did it (fish) suddenly appear out of water?' (*Did it go BILU?*)
b. l-á-rɪ phátsi? 'did lightening flash?' (*Did it go PHATSI?*)
c. a-rɪ tshíki? 'is it cold?' (*Did it go TSHIKI?*)
d. ányí a-rɪ bítsɪ 'who left in a hurry?' (*Who went BITSI?*)

Questions also override the final lengthening in paused lists:

(29) Yes-No Q >> List (Ø >> FL)

a. a-bal-a ri-nama . . . ri-nawá . . . kana ri-nárí?
'has he just counted meats. . . beans. . . or buffalos?'
b. a-bɔ́n-á lʊ-rʊli . . . malílɪ . . . kana mʊ-rírí?
'has he just seen dust. . . rubbish. . . or hair?'

The examples in (30) and (31) show that the final devoicing of ideophones overrides the Ø of the hortative and the final lengthening of paused lists:

(30) Ideo >> Hort (FD >> Ø)
á ba-rɪ bítsɪ̥ 'may they leave in a hurry!'

(31) Ideo >> list (FD >> FL)
y-á-rɪ bílʊ̥ . . . bítsɪ̥ . . . phátsi̥ . . . tshíki̥ 'it suddenly appeared out of water, in a hurry, flash of lightning, cold'

The fact that interrogative Ø overrides the FD of ideophones, but the FD of ideophones overrides the Ø of hortatives and paused lists provides the motivation for splitting the Ø intonation into different positions on the hierarchy. Finally, the sentences in (32) show that the Ø of both WH-questions and imperatives block the final lengthening of paused lists:

(32) WH-Q, Imper >> List (Ø > FL)
   a. ányí a-bal-a ri-nama . . . ri-nawá . . . lí ri-nárí
      'who is counting meats. . . beans. . . and buffalos?'
   b. bal-á ri-nama . . . ri-nawá . . . lí ri-nárí
      'count meats. . . beans. . . and buffalos!'

The examples in (28)–(32) and the resulting hierarchy in (27) illustrate what we might refer to as the pragmatically unmarked way of resolving conflicts between the different utterance types and their intonations. There is, however, evidence that speakers have further options available to them if they want to place a different emphasis on an utterance. Before going into this, it must be reiterated that everything that has been shown up to this point represents a neutral or non-emphatic realization, whether of a specific utterance type or of a conflict between utterance types. The additional options we are about to illustrate were never directly elicited, i.e. when translating an English utterance into Shegkalagari. Rather, it was only when we systematically assigned alternative intonations to different utterance types to see if they were interpretable that we discovered other possibilities.

A major complication with respect to the hierarchy in (27) is a marked highlighting process, possibly paralinguistic, which we term 'emphatic' PLL. While declarative PLL is pragmatically unmarked, Emph PLL is highly marked and can be assigned to any utterance type except yes-no questions, with seemingly contradictory results such as those in (33).

(33) Emph PLL can
   a. make WH-Qs, imperatives, and hortatives seem either like statements or more insistent
   b. emphasize or de-emphasize the effect of such non-declarative speech acts
   c. clarify what was said, often repeating or rewording when someone has not understood
   d. provide some kind of emphasis, but not necessarily on the last word or even the last constituent
   e. be often subtle, never obligatory, perhaps 'attitudinal' in the sense of Bolinger (1978: 484)

Consider the case of WH-words. First, as seen in (34), it is not surprising that they undergo PLL in citation form:

(34)	citation form	as question		citation form	as question	
	â:nyɪ	ányí	'who'	lî:ŋ̀	líŋ́	'when'
	ʃi:ŋ́	ʃiŋ́	'what'	ʒwâ:ŋ̀	ʒwáŋ́	'how'
	qâ:ɪ	qáí	'where'	qa xʊri:ŋ́	qa xʊriŋ́	'why'

In an elicitation session, if one asks a speaker how to say 'who', the answer has to be *â:nyɪ*, since the form is not a question, but rather a declarative citation form. (If asking for the one-word question utterance 'who?' the form would of course be *ányí*.) The same is observed in (35a), where the WH-word is used in a contrastive declarative utterance:

(35) a. ʊ́-ráy-á qâ:ɪ 'you mean WHERE'
b. ʊ́-ráy-á qáí 'do you mean WHERE?'

The absence of PLL in (35b) unambiguously establishes the utterance as a question. (Although incomplete, (35b) could also be taken to mean 'where do you mean?') Since the WH words in (34) and (35a) do not occur in Wh-questions, they escape interrogative Ø-assignment in (27) and instead filter down to receive declarative PLL.

The situation in (36), however, is quite different.

(36) a. w-á qáí 'where are you going?':
b. w-á qâ:ɪ 'where are you going?'
c. yó ꜜíye í-ꜜbɔ́n-á Mʊnaká kí ꜜányí (~ꜜâ:nyɪ)
DEM PAST s/he-saw Monaka is who
'the one who saw Monaka is who?'

While the normal WH-question is with Ø, as in (36a), (36b) can be used to repeat the question, for insistence, or 'just emphasis'. In (36b) the most immediate interpretation is that the speaker is being very insistent: he or she really wants to be responded to! The same interpretation occurs when PLL applies in (36c): 'The one who saw Monaka is WHO? Tell me!'

While EMPH PLL intensifies the illocutionary force of a WH-question, it seems to have an attenuating effect on imperatives and hortatives. As seen in (37), the unmarked Ø forms are interpreted as commands, while the forms with EMPH PLL seem rather to be suggestions:

(37) a. ḿ-bʊ́ʒ-ɛ́ 'ask him!'
ḿ-bʊ̂:ʒ-ɛ 'what you can do is ask him' (that's what I suggest)
3sg-ask-INFL
b. ʃɪ́-ǹ-tʃwɛ́l-ɛ́ 'don't tell me!' (= an instruction)
ʃɪ́-ǹ-tʃwɛ̂:l-ɛ 'you shouldn't tell me!' (= a statement)
NEG-1sg-tell-INFL

c. á ꜜbá-bál-ɛ ri-nárí 'let them count the buffalos!' (= stronger, a command)
á ꜜbá-bál-ɛ ri-nâ:rɪ 'they should count the buffalos!' (= weaker, a suggestion)
COMP-3pl-count-INFL buffalos

Other cases were found to have just the opposite effect. Thus, the imperative/hortative forms with PLL in (38) mark insistence or have a strong finality effect ('and that's that!'):

(38) a. ʃí-gy-ɛ́ 'eat it!' (= normal)
b. ʃî:-gy-ɛ 'eat it!' (= stronger)
c. á kɪ-gy-ɛ kʊ́kʊ́ 'let me eat the chicken!' (= weaker)
d. á kɪ-gy-ɛ kʊ̂:kʊ 'let me eat the chicken' (= stronger)
COMP-1sg-eat-INFL chicken

While (38a) is the normal imperative, the PLL form in (38b) might be translated 'eat it or else!' or 'eat it already!' (with the speaker showing impatience). The normal hortative in (38c) is weak enough to be interpreted as asking permission ('may I eat the chicken?'), while the PLL form in (38d) expresses finality ('that's what I'm going to do!') and does not expect an answer.

In still other cases, the effect of emphatic PLL is not clear, other than adding a vague sense of emphasis:

(39) a. balá ꜜrí-nárí (. . .) ísɪŋ́ ri-kú 'count buffalos, not sheep!'
b. balá ꜜrí-nâ:rɪ . . . ísɪŋ́ ri:-kú (idem.)

Finally, it should be noted that EMPH PLL can override anything except a yes-no question (which would then become a statement):

(40) a. >> IDEO FD : y-á-rɪbílʊ̥ ~ y-á-rɪ bí:lʊ
'it suddenly appeared out of water'
b. >> monosyllabic word Ø : bá-bal-a ʒé ~ bá-bal-a: ʒé
'they are counting these'
(recall the failure of PL to apply in (24))
c. >> WH-Q Ø >> IDEO FD : ányí a-rɪ bí:tsɪ
'who left in a hurry?' (insisting)

In all of the above examples with EMPH PLL the basic utterance type is recoverable from the structure: presence of a WH-word, absence of a subject in imperatives, presence of *á* in hortatives and exclamatives, the verb *-rɪ* 'say' with ideophones, etc. Since yes-no questions are marked exclusively by Ø intonation, if the penultimate vowel were to be lengthened the result would be a statement, not a question. Emph PLL thus may not occur on a yes-no question.

While the hierarchy in (27) can be modified to accommodate Emph PLL, Yes-No and WH-Qs would have to be split up, as in (41).

(41)	Yes-No Q	>>	Emph	>>	Wh-Q
	Ø		PLL		Ø

Since we have argued that Emph PLL cannot override the Ø of a yes-no question for reasons of recoverability, it is not clear that the hierarchical approach in (41) is the right way to go. An alternative interpretation is that Emph PLL represents a separate (paralinguistic?) dimension, perhaps like Bolinger's (1972: 644) characterization of 'accent': 'The distribution of sentence accents [in English] is not determined by syntactic structures but by semantic and emotional highlighting.' And so it appears to be with Emph PLL.

There is in fact good reason to view the original hierarchy in (27) as representing the 'normal' or 'expected' relationships between the different intonations, which can, however, be modified in marked situations. Quite late in our study we discovered the following minimal triplet concerning ideophones in WH-questions:

(42) a. ányí a-rɪ bítsɪ 'who left in a hurry?'
b. ányí a-rɪ bítsɪ̥ 'who left in a HURRY?'
c. ányí a-rɪ bí:tsɪ 'WHO LEFT IN A HURRY?'

As we have said, (42a) represents the normal or expected form, where Wh-Q Ø overrides the FD of the ideophone. Although the other two possibilities are quite unusual, (42b) might be uttered if the speaker wanted to bring special emphasis to leaving in a hurry, perhaps contrasting *bítsɪ̥* with another ideophone. In (42c), with Emph PLL, emphasis is on the whole question, as when the speaker, perhaps with exasperation, is insisting that s/he be responded to. What's important is that speakers do have some choice in effecting different pragmatic overrides violating the hierarchy in (27).

The final question in this section is how to account for the variation in the meanings of Emph PLL. As an introduction to what might be going on, consider the utterances in (43).

(43) a. íye bá-m-bʊ́ʃ-a xʊ-rɪ íye a-bɔ́n-a â:nyɪ 'they asked him who he saw'
b. íye bá-m-bʊ́ʃ-a xʊ-rɪ íye a-bɔ́n-a ányí
past 3pl-3sg-ask to say past 3sg-see who
'did they ask him who he saw?'

As can be observed, indirect questions take PLL or Ø according to the nature of the higher-clause: (43a) is a statement, while (43b) is a question. What we would like to propose is that, like indirect questions, EMPH implies an abstract declarative higher clause, hence PLL. As indicated in (44), the unexpressed higher clause (in parentheses) may have either an emphatic or attenuating effect:

(44)	a.	WH-Q	Ø	:	Where are you going?	
			PLL	:	(I am asking you again) where you are going?	
	b.	IMPER	Ø	:	Ask him!	
			PLL	:	(What I suggest is) ask him!	(= weaker)
			PLL	:	(Again, I'm telling you to) ask him!	(= stronger)
	c.	HORT	Ø	:	Let them count the buffalos!	
			PLL	:	(What I suggest is that) they count the buffalos! (= weaker)	
			PLL	:	(What I suggest is that) they count the buffalos! (= stronger)	

In other words, EMPH PLL may be paralinguistic (Ladd 1996) and attitudinal (Bolinger 1978), outside the structural system and subject to cultural norms. (We note, for example, that one cannot use an imperative + EMPH PLL if speaking to an older person.) As such, it is hard to pin down exclusive or fixed meanings. We thus arrive at a view of EMPH PLL much like Bolinger's characterization of pitch: 'The picture is clouded in a number of ways. The meanings conveyed by pitch are attitudinal, and attitudes are notoriously subject to distortion and inhibition.. . .' (Bolinger 1978: 515). Thus, to paraphrase Bolinger (1972), EMPH PLL may be 'predictable' if you are a mind-reader.

Before leaving this section, we would like to make one more point: While we believe the above characterization of the different intonations in Shekgalagari to be accurate, there is undoubtedly much more to be said. One issue we have not dealt with is phrasing. This is another area where there is an expected realization, but also some choice. Consider, for example, the utterances in (45), which concern the marking of right-dislocations:

(45)	a.	ba-rímʊ́ bá-bál-a ri-nâ:rɪ gods they-count buffalos	'the gods have just counted the buffalos'
	b.	bá-rí-bá:l-a they-them-count	'they have just counted them'
	c.	[bá-rí-bál-a, ba-rímʊ́, ꜜrí-nâ:rɪ]	'they have just counted them,
	d.	[bá-rí-bál-a, ba-rímʊ́:] [ꜜrí-nâ:rɪ]	the gods, the buffalos'
	e.	[bá-rí-bá:l-a] [ba-rímʊ́:] [ꜜrí-nâ:rɪ]	
	f.	[bá-rí-bá:l-a] [ba-rî:mʊ] [ri-nâ:rɪ] they-them-count gods buffalos	(= 'emphasis' (EMPH))

(45a) shows the pervasive SVO structure of Shekgalagari. When the class 10 object *rinâ:ri* 'buffalos' is pronominalized in (45b), the prefix *-rí-* occurs in its place. The utterance in (45c) shows the 'normal' way of expressing right-dislocations. As seen, there is no pause, and therefore only the last word is marked by PLL. In (45d-f) the fully bracketed nouns indicate that there is a pause before them. In (45d, e) we see that the recapitulated subject *ba-rím ʊ́*: 'the gods' undergoes FL, which we have heretofore identified with paused lists. In addition, the prepausal verb undergoes PLL in (45e). In (45f), which sounds very emphatic, each of the three pause groups is marked by PLL, as if to say, 'I'm telling you they COUNTED them, the GODS, the BUFFALOS.' While the lack of pauses in (45c) is the most natural realization of right-dislocations, the above realizations give some idea of the range of variation that is potentially available to Shekgalagari speakers. No doubt further investigation will turn up more subtleties and clarification of the relation between intonation marking and phrasing.

5 Summary and Conclusion

In the preceding sections we have seen that Shekgalagari, a tone language, is rich in intonational options. Except for PLL, which includes a L% feature, intonation is not tonal, but rather involves penultimate lengthening, final lengthening, final devoicing, or none of the above. Returning to the three strategies a tone system may adopt for dealing with intonation (accommodation, submission, avoidance), Shekgalagari's response seems best characterized as accommodation: None of the intonations merges anything from the lexical phonology. Thus, while the L% of PLL has a striking effect on tone, HL:-H and LꜜL:-L unambigously correspond to medical H-H and L-L. Similarly, the length from PLL and FL does not cause merger, since there is no underlying lexical contrast in the language. (We assume that the effects of Proto-Bantu long vowels on tone indicated in (7) is not best analyzed by setting up an abstract, underlying vowel length contrast.) In fact, it has long been observed that Bantu languages which have lost the historical vowel length contrast are more likely to have penultimate lengthening: '...many Bantu languages have an H and L tone with a superimposed penultimate accent. This accent may cause vowel lengthening (especially if the vowel length contrast of Proto-Bantu has been lost), or it may affect the tone of the penultimate syllable' (Hyman 1978:14). Finally, there is no loss of lexical information when ideophones undergo final devoicing. It can thus be said that Shekgalagari has found ways to express different intonations without infringing on the prosodic properties of the word-level phonology.

Several of the properties of the Shekgalagari intonational phonology are of typological interest: (i) the specific utterance types that are distinguished (e.g. imperative, ideophone); (ii) the non-tonal means by which utterance types are distinguished (lengthening, devoicing, absence of marking); (iii) the hierarchization of the intonational functions (which is reminiscent of competing tonal assignments in inflectional morphology); (iv) the pragmatically marked nature of Ø intonation—which can override the others. The system is also of diachronic interest in the sense that the PLL, FD and FL intonations are not likely relics of old particles that been lost, but are more probably due to independent phonologizations.

The Shekgalagari system naturally raises the question of what the full range of intonational features is. Besides pitch, length, and devoicing, attested in Shekgalagari, both breathy and creaky phonations as well as the laryngeal segments -*h* and -ʔ have been claimed to mark intonation. For example, final glottal stop marks imperatives in Lahu (Tibeto-Burman; Thailand, China, Myanmar) (Matisoff 1973: 353), questions in Kaingang (Macro-Ge; Brazil) (Wiesemann 1972, Wetzels 2008), and negatives in Dagbani (Niger-Congo, Gur; Ghana) (Hyman 1989). Perhaps other prosodic features such as nasality may also be exploited for intonation (cf. the 'nasal pause' phenomenon in Amazonia (Aikhenvald 1996: 511-512)). Whether suprasegmental segments such as -*k*, -*s*, or -*m* or full syllables can mark intonation has been questioned (Hyman 1989), although Aikhenvald (1998: 410) reports the case of -*hĩ*or *h* + a V copy with nasalization in Warekena (Arawakan; Venezuela): 'A morpheme -*hṽ* 'pausal marker' is inserted at the end of a phonological word or a phonological phrase.' We have cited Matisoff's (1994: 118) comment about the equivalence of particles to intonation marking, a point also taken up by Ladd (1996) and in earlier work of my own: '…besides the parallel in function, there may be important structural similarities between boundary tones and particles. In fact, the difference may be simply that the former lack segmental content, while the latter do not' (Hyman 1990: 123). Thus, paralleling intonational marking in other languages, 'interrogative, exclamatory, imperative, emphatic, and doubt meanings in Capanahua are represented in the base component by features that are spelled out as segments (morphemes)' (Loos 1969: 211). (Cf. paralinguistic [+irritation] marking in Capanahua: 'Anger and irritation are not given morphemic shape in the string, but are expressed... by nasalization of the whole sentence'.) This raises the question of whether prosodic 'intonemes' are morphemes (Hyman 1990), and if so, whether they should be identified as clitics or phrasal affixes in the sense of Anderson (1992, Chap. 8).

This brings us to the final question: What are the necessary definitional properties of intonation? It seems there are at least three possibilities in determining what should vs. should not be considered 'intonation': One might restrict intonation to certain specific realizations (pitch, duration etc.). Alternatively, one might delimit intonation on the basis of a restricted set of functions (declarative, interrogative etc.). A final possibility is that intonation might be identified in terms of its domain or place in a grammar. In this last case, we might say that anything that originates at the intonational phrase or utterance level, or within the 'Phonetic Form' module of government-binding theory, is by definition 'intonation'. In this last approach it would not matter if the mark were a feature, a mora, a segment, or a fuller 'particle'. The equivalence would be defined by the place at which the so-defined intonation enters the grammar. Conversely, a feature which has to be present earlier in the phonology would not be intonational, nor would a particle which has to be present in the syntax. We believe that this kind of approach is likely to be the most revealing in determining what is vs. is not intonation.

APPENDIX: Monosyllabic Words and PLL

As was seen in (24), declarative PLL does not apply when the last word of the utterance is monosyllabic, a property which is thus far limited to Shekgalagari among the Bantu languages for which we have information. Here we consider a few more facts in order to determine how this fact might be account for.

First, it should be noted that monosyllabic words are very limited in Shekgalagari. Among the ones we have identified by independent criteria are the following:

(46) a. monosyllabic verbs in the imperative: *k-á* 'mention!', *gy-á* 'eat!'
b. demonstratives: *ʃé* 'this (one)' (cl. 7), *ʒé* 'these' (cl. 10) etc.
c. adverbs: *ʒwá* 'in this way', *thɛ́* 'really'
d. the preposition *qá* 'with', which, however, cannot occur finally

As seen, all of the above monosyllabic words have /H/ tone. Monosyllabic imperatives have a bisyllabic variant, which can occur with EMPH PLL: *ɪː-k-á* 'mention!', *ɪː-gy-á* 'eat!'. Similarly, although monosyllabic words block declarative PLL, EMPH PLL may assign length to the final vowel of the preceding word:

(47)	a.	a-bal-a qá ʃe	~	a-bal-a qá: ʃe	'he has just counted with this'
	b.	a-rí-bál-a ʒwá	~	a-rí-bál-a: ʒwá	'he has just counted them like this'

(In (47a) /qá + ʃé/ 'with this' becomes *qá ʃe* by a rule discussed by Crane (2008, 2009a, b).)

But should utterance-level intonation, here PLL, be allowed to have access to word boundaries? If yes, monosyllabic words can block PLL by virtue of not having a penultimate syllable. If no, an alternative is needed to avoid direct reference to word boundaries.

While we suspect that intonation can know where the word boundaries are, if it were necessary to exclude them from intonational implementation, the following metrical solution would work:

(48) a. construct a trochaic foot over the last two syllables of each word
b. in case the last word is monosyllabic, the trochaic foot will have only one syllable
c. declarative PLL specifically targets the nucleus of the penultimate syllable (vowel or syllabic nasal) of the last foot of an utterance or pause-marked intonational phrase (IP)
d. Emph essentially encliticizes an IP-final monosyllabic word in which case PLL is free to target the nucleus of the penultimate syllable across the word boundary.

Acknowledgments Besides the TIE3 conference, the first author presented this paper as colloquia at the University of California, Berkeley and M.I.T. We are grateful for the comments we received at all three presentations and individually, particularly from Carlos Gussenhoven and Keith Johnson, as well as from the editors and two anonymous reviewers.

References

Aikhenvald, Alexandra Y. 1996. Words, phrases, pauses, and boundaries: Evidence from South American Indian Languages. *Studies in Language* 20:487–517.
Aikhenvald, Alexandra Y. 1998. Warekena. In Desmond C. Derbyshire and Geoffrey K. Pullum (eds.) *Handbook of Amazonian languages*, vol. 4, 225–439. Berlin and New York: Mouton de Gruyter.
Anderson, Stephen R. 1992. *A-Morphous Morphology*. Cambridge University Press.
Bearth, Thomas, and Christa Link. 1980. The tone puzzle of Wobe. *Studies in African Linguistics* 11:147–207.
Bearth, Thomas, and Hugo Zemp. 1967. The phonology Dan (Santa). *Journal of African Languages* 6:9–29.
Bolinger, Dwight. 1972. Accent is predictable (if you're a mind-reader). Language 48: 633–644.
Bolinger, Dwight. 1978. Intonation across languages. In Joseph H. Greenberg (ed.) *Universals of Human Language*, vol. 2, 471–524. Stanford University Press.
Crane, Thera. 2008. Predicting downstep in Shekgalagari. Ms. University of California, Berkeley.
Crane, Thera. 2009a. Evaluating approaches to downstep in Shekgalagari. Paper presented at the Annual Meeting of the Linguistic Society of America, San Francisco, January 9, 2008.
Crane, Thera. 2009b. *Tone, Aspect and Mood in Shekgalagari*. Phonology Laboratory Annual Report, 2009. Berkeley: University of California, Berkeley. http://linguistics.berkeley.edu/phonlab/annual_report/documents/2009/CraneShekgTAMDraftFeb2009forPhonLab.pdf
Dickens, Patrick. 1984. Qhalaxarzi verb tone classes. *African Studies* 43: 109–118.
Dickens, Patrick. 1986a. *Qhalaxarzi Phonology*. M.A. Dissertation, University of Witwatersrand.
Dickens, Patrick. 1986b. Tone in Qhalaxarzi main verb constructions. *South African Journal of African Languages* 6: 67–70.
Gralow, Frances L. 1985. Coreguaje: Tone, stress and intonation. In Ruth M. Brend (ed.) *From Phonology to DisCourse: Studies in Six Colombian Languages*, 3–11. Lang. Data, Amerindian Series No. 9. Dallas: SIL.
Gussenhoven, Carlos. 2004. *The Phonology of Tone and Intonation*. Cambridge University Press.

Guthrie, Malcolm. 1967-71. Comparative Bantu, vols. 1–4. Farnborough: Gregg International Publishers.

Hirst, Daniel, and Albert Di Cristo. 1998. A survey of intonation systems. In Daniel Hirst and Albert Di Cristo (eds.) *Intonation systems*, 1–44. Cambridge University Press.

Hockett, Charles F. 1963. The problem of universals in language. In Joseph H. Greenberg (ed.) *Universals of Language*, 1–29. Cambridge, MA: M.I.T. Press.

Hyman, Larry M. 1978. Tone and/or accent. In Donna Jo Napoli (ed.) *Elements of Tone, Stress and Intonation*, 1–20. Washington: Georgetown University Press.

Hyman, Larry M. 1989. The phonology of final glottal stops. In *Proceedings of WECOL 1988*, 113–130.

Hyman, Larry M. 1990. Boundary tonology and the prosodic hierarchy. In Sharon Inkelas and Draga Zec (eds.) *The Phonology-Syntax Connection*, 109–125. Chicago and London: Univerity of Chicago Press.

Hyman, Larry M. 2009. Penultimate lengthening in Bantu. Submitted to Festschrift for Johanna Nichols. http://linguistics.berkeley.edu/phonlab/annual_report/documents/2009/Hyman_Penult_Length_PLAR.pdf

Hyman, Larry M., and Joyce T. Mathangwane. 1998. Tonal domains and depressor consonants in Ikalanga. In Larry M. Hyman Hyman and Charles W. Kisseberth (eds.) *Theoretical Aspects of Bantu Tone*, 195–229. Stanford: C.S.L.I.

Janson, Tore. 1995. The status, history and future of Sekgalagadi. In A. Traill, R. Vossen and M. Biesele (eds.) *The Complete Linguist: Papers in Memory of Patrick J. Dickens*, 399–406. Cologne: Rüdiger Koppe Verlag.

Krüger, C. J. H., and J. A. du Plessis. 1977. *Die Kgalagadi Dialekte van Botswana*. Potchefstroomse Universiteit vir Christelike Hoer Onderwys.

Ladd. D. Robert. 1980. *The Structure of Intonational Meaning*. Bloomington: Indiana University Press.

Ladd, D. Robert. 1996. *Intonational phonology*. Cambridge University Press.

Loos, Eugene E. 1969. *The Phonology of Capanahua and its Grammatical Basis*. Norman: SIL and University of Oklahoma.

Lukusa, Stephen T. M., and Kemmonye C. Monaka. 2008. *Shekgalagari Grammar*. Casas Book Series 47. http://www.casas.co.za/Publications.aspx?SCATID = 3.

Mathangwane, Joyce T. 1999. *Ikalanga Phonetics and Phonology: A Synchronic and Diachronic Study*. Stanford Monographs on African Languages. Stanford: CSLI.

Matisoff, James A. 1973. *The Grammar or Lahu*. Berkeley: University of California Press.

Matisoff, James A. 1994. Tone, intonation, and sound symbolism in Lahu: loading the syllable canon. In Leanne Hinton, Johanna Nichols and John J. Ohala (eds.) *Sound Symbolism*, 115–129. Cambridge University Press.

Monaka, Kemmonye C. 2005a. Shekgalagari laryngeal contrasts: the plosives. *South African Journal of African Languages* 25: 243–257.

Monaka, Kemmonye C. 2005b. VOT in Shekgalagari stops. *Lwati* 2: 24–42.

Monaka, Kemmonye C. In preparation. Trilingual dictionary of Shekgalagari, Setswana and English (6600 + entries).

Neumann, Sabine. 1999. *The Locative Class in Shengologa (Kgalagadi)*. P. Lang: Frankfurt am Main.

Pike, Eunice V. 1951. Tonemic-intonemic correlation in Mazahua (Otomi). *International Journal of American Linguistics* 17: 37–41. Reprinted in *Studies in Tone and Intonation*, ed. Ruth M. Brend, 100–107. Basel: S. Karger, 1975.

RETENG (The Multicultural Coalition of Botswana). 2006. Alternative Report Submitted to the UN Committee on the Elimination of All forms of Racial Discrimination (CERD). Gabarone, Botswana.

Singler, John Victor. 1984. On the underlying representation of contour tones in Wobe. *Studies in African Linguistics* 15: 59–75.

Vydrine, Valentin, and Mongnan Alphonse Kességbeu. 2008. *Dictionnaire dan-français (dan de l'Est*). Nestor-Istoria: St. Petersbourg.

Wetzels, W. Leo. 2008. Word Prosody and the distribution of Oral/Nasal Contour Consonants in Kaingang. Talk given at the Laboratoire de Phonetique et Phonologie (Paris 3), May 23, 2008.

Wiesemann, Ursula. 1972. *Die phonologische und grammatisch e Struktur der Kaingang Sprache*. The Hague: Mouton.

Subject Index

S. Frota et al. (eds.), *Prosodic Categories: Production, Perception and Comprehension*, Studies in Natural Language and Linguistic Theory 82,
DOI 10.1007/978-94-007-0137-3, © Springer Science + Business Media B.V. 2011