Tone sandhi, prosodic phrasing, and focus marking in Wenzhou Chinese

Published by
LOT
Trans 10
3512 JK Utrecht
The Netherlands

phone: +31 30 253 6006

e-mail: lot@uu.nl
http://www.lotschool.nl

Cover illustration: Franziska Scholz

ISBN: 978-94-6093-089-8
NUR 616

Tone sandhi, prosodic phrasing, and focus marking in Wenzhou Chinese

Proefschrift

ter verkrijging van
de graad van Doctor aan de Universiteit Leiden,
op gezag van Rector Magnificus prof.mr. P.F. van der Heijden,
volgens besluit van het College voor Promoties

te verdedigen op donderdag 18 oktober 2012
klokke 11.15 uur

door

Franziska Scholz

geboren te Berlijn, Duitsland
in 1983

Promotiecommissie

Promotores:	Prof.dr. L.L.-S. Cheng Prof.dr. V.J.J.P. van Heuven
Co-promotor:	dr. Y. Chen
Commissieleden:	Prof.dr. R. You (Fudan University, Shanghai) dr. L. Downing (ZAS, Berlin)

The research reported in this thesis was conducted in the context of the project "An experimental approach to the interaction of tone sandhi and focus expression in six dialects of Chinese" (principal investigator: dr. Yiya Chen). This project was funded by the Netherlands Organisation for Scientific Research (NWO) with a VIDI grant (061084338).

Meinen Eltern gewidmet

Contents

Acknowledgements

As most theses, this thesis would not have been possible without the support and encouragement of many many people. Leiden regulations do not allow me to name and thank the three people who have most contributed to the successful completion of my project. Each of them has shared their unique expertise with me, and they have complemented each other in the most fruitful way.

Secondly, I would like to thank the members of my reading committee for devoting their time and attention to my research. Prof. Rujie You shared his insights on Wenzhou Chinese and tone sandhi, and dr. Laura Downing contributed on the topics of prosodic phrasing and information structure. I also thank the members of my defense committee for taking part in my ceremony.

Next in line, I would like to thank my two paranimfen, Elizabeth and Judith. Both have helped me in more ways than I can count, shared the joys and woes of PhD life with me, and kept me from losing my mind any more than I did. I consider myself very fortunate for having them by my side during my defense, and for having made great friends in them.

One of the most important tasks for a PhD student is to ask questions, and I was in the lucky position to never have to go far in order to find someone to ask. For their patience with me, I thank my many office mates through the years who have made the cold Lipsius 1.04 a warmer place: Amanda, Daan, Hongyan, Jia Yuan, Junru, Jurriaan, Laurent, Rongjia, Sandra, and Willemijn. Particular thanks go out to Daan for frequent help with random Dutch grammar questions, and to Rongjia and Junru for helping me when I was driven close to desperation by Chinese characters.

Special mention is also due for the senior staff in the Leiden Phonetics Lab. Throughout the course of my PhD, Willemijn, Rachel, and Jos have always been willing to help me and, most of all, to explain things to me so I could do them by myself the next time. All three of them have gone far above and beyond what could be expected of “just” colleagues who weren’t actually part of my supervision team, and they have become dear friends. Thank you!

Outside of the Phonetics Lab, LUCL has been a great place to meet researchers from all different corners of the globe, and I feel enriched by the experiences that they shared with me. Special thanks for frequent conversations

and discussions as well as for occasional moral support go to Björn, Kathrin, Linda, and Monica.

Learning Chinese on your own is a frustrating and unthankful endeavor that no one in their right mind should ever set out to undertake. That's one side of it. For helping me see the other side of it, namely the beauty of the Chinese language and culture and the intellectual growth that is the fruit of its study, I would like to thank my Chinese language partners. Special thanks go to Qian Kefei, for having ignited the first spark of interest for Chinese in me, and for being the single most enthusiastic and positive person I know. Here in Leiden, it has been a pleasure to meet up every week with Chen Lijun and Luo Yixue for Chinese lessons, and to teach them a thing or two about German in return.

Studying a (comparatively) rare dialect of Chinese without the patient and repeated help of native speakers is impossible. I was in the very lucky position to not only be able to tap into a pool of enthusiastic speakers in Wenzhou city, but also to find great informants here in the Netherlands. First and foremost, thanks to Wang Zheng for always being on stand-by to help me figure out the intricacies of his native dialect, and for patiently recording one boring task after another without ever tiring. You're a star!

Thank you also to Chen Yinyin for not only being an enthusiastic informant during her stay in the Netherlands, but also for welcoming me on my field trips to her home town. Without her organizational talent and persuasion powers, my second fieldtrip would have gone wrong on so many levels! Thanks also to Lin Zhe for a short, but very fruitful recording cooperation in Leiden. Lastly, thank you to all my recording subjects in Wenzhou city, who were excited to help me complete my studies, and made it through my boring experiments without falling asleep (all but one...). You made this possible!

All of the above is about work, but all of life is not. My last but certainly not least set of *thank you*'s goes to my friends and family. Thank you to Konrad, for being my best friend and fellow e-mail-as-novel writer (you realize that our collected e-mails of the last four years exceed the volume of this thesis threefold?), for bestowing me with his wisdom and advice, and for setting my head straight when I need it. Thank you to Martina, for always being there, and for reminding me where I come from. Thank you to Lina, for sharing her adventures with me, and for being my little sister in spirit. Thank you to Kilu, for being the most awesome unobtainable role model and a great inspiration. You all rock, and I couldn't have made it without you!

And finally, thank you to my family, for having been my safe haven, and for enabling me to do what I did. Particularly, thank you to Elli, Micha, Uta

and Thomas for their continued interest in my arcane work. And most of all, thank you to my parents, for raising me up to be curious and thirsty for knowledge, for letting me go out in the world and discover it, and for enthusiastically sharing in what I found. *Mama, Papa, ihr habt den Grundstein gelegt für alles, was ich in meinem Leben erreicht habe. Danke für eure Unterstützung, euer Vertrauen, und eure Liebe.*

My last thank you goes to Dragan, for having found me as I found him (with some help from Elizabeth), and for having been in my corner ever since. Thank you for sticking it out with me through all the occasional highs and frequent lows, for making me laugh when I needed it and being my shoulder to cry on when I needed it, and for being my rock in stormy seas. I'm looking forward to spending the rest of my life with you.

Franziska

Abbreviations and symbols

ADV	adverb
ASP	aspect marker
BF	broad focus
C	consonant
CL	classifier
CP	complementizer phrase
DP	determiner phrase
F_0	fundamental frequency
H	high register
H	local F_0 maximum
Hz	Hertz
IP	inflection phrase
L	low register
L	local F_0 minimum
N	noun
NP	noun phrase
NUM	numeral
O	object
OBJ	object marker
OF	object focus
φ	phonological phrase
Q	question particle
S	subject
SF	subject focus
st	semitones
σ	syllable
V	vowel
V	verb
VP	verb phrase
VPF	VP focus
ω	prosodic word
*	prosodic head

Chapter 1

Introduction

This thesis looks at data from Wenzhou Chinese to address several research questions that lie at the intersection of three areas of linguistic interest: (i) tonal realization and tone sandhi, (ii) prosodic phrasing, and (iii) information structure, in particular focus. The following sections provide a short introduction into each of these areas, and to the connections between them. After that, the goal of this thesis as well as the individual research questions will be laid out, and related to the respective chapters of the thesis that address them.

1.1 Issues addressed in this thesis

1.1.1 Tonal realization and tone sandhi

Like virtually all Sinitic dialects, the dialect of Wenzhou uses *tone*, or complex F_0 (fundamental frequency) modulations, to distinguish the lexical meanings of segmentally identical strings of sound. While the research on tone has a long tradition especially in Chinese linguistics, the recent availability of modern research tools has resulted in a massive increase in experimental research on tones and tonal changes in the last two decades.

A topic that has been addressed in many of these publications has to do with the exact implementation of lexical tones in different environments and contexts. Lexical tones have been found to be affected in their implementation by neighboring tones (Shen & Lin 1991; Xu 1994, 1997; Xu & Wang 2001), the function of the sentence (e.g. declarative or interrogative) of the sentence they appear in (Jiang & Chen 2011; Liu & Xu 2005; Shen 1990a; Shih 2004), the strength of the tone-bearing unit (Chen & Xu 2006), the discourse context (Chen et al. 2009; Chen 2009, 2010; Chen & Gussenhoven 2008; Wang & Xu 2011; Xu 1999; Xu et al. 2012), and their prosodic environment (Brooke et al. 2009; Cao & Zheng 2006; Chow 2006; Pan & Tai 2006; Peng 1997; Shen 1992b; Shih 1997; Xu & Wang 2009; Yang & Wang 2002).

This is not surprising when considering the multiple functions of F_0 in speech, and the physiological constraints that govern tonal implementation. The F_0 contours that are the main characteristics of lexical tones obey certain

restrictions, such as a maximum speed with which pitch changes can be implemented (Xu & Sun 2002), and influence from vowels (Whalen & Levitt 1995) and consonants (Cao & Maddieson 1992; Chen 2011) (see Xu, Y. 2004 for an overview). Additionally, a common finding in Chinese dialects is that speakers employ certain types of changes to the tonal implementation for communicative purposes, for example to mark a number of adjacent tones as lexically or semantically associated. This type of tone change is usually referred to as *tone sandhi* (变调 'biàndiào' in Chinese) (Chen 2000).

The exact line of distinction between *tone sandhi* and *tonal coarticulation* has been subject to debate. In most research articles, it is (often implicitly) assumed that tone sandhi involves a phonological and categorical process, whereas tonal coarticulation is phonetic (see e.g. Zhang & Liu 2011, passim in Zhang & Lai 2010). Analogously to the distinction between assimilation and coarticulation in segments, Shen (1992a) proposed to differentiate tone sandhi from tonal coarticulation according to three criteria: differences in mechanism (tone sandhi = dependent on "language-specific morphophonemic constraints", tonal coarticulation = "attributed to language-independent biomechanical constraints"), process (tone sandhi = can be assimilatory or dissimilatory, tonal coarticulation = is always assimilatory), and identity (tone sandhi = changes tonal identity, tonal coarticulation = preserves tonal identity).

However, the distinction is criticized in Chen 2000, who argues that even phonetic tone change can be "grammatically controlled", assimilatory or dissimilatory, and that tone sandhi can be allotonic and does not necessarily have to be category-changing. Nonetheless, also Chen excludes "low-level phonetic coarticulation effects" from his investigation of tone sandhi, since he is relying on impressionistic data which is limited by perceptibility. In that sense, it is unclear how Chen's distinction could be adapted to the current study, which aims to investigate both tone sandhi and tonal coarticulation with instrumental methods. For the purpose of nomenclature, this thesis will reserve the term *tone sandhi* to refer to the particular type of tone change that affects disyllabic lexical words in Wenzhou, and talk about *tonal coarticulation* in other contexts where lexical tones are realized with tonal contours that are influenced by neighboring tones.

1.1.2 Prosodic phrasing

As will become evident from the data presented in the following chapters, tone sandhi in Wenzhou Chinese applies to disyllabic constituents, and its application

serves to mark these disyllabic constituents as lexicalized compounds. In other dialects of Chinese, however, the tone sandhi domain may not be so clearly demarcated. Often, accounts of tone sandhi have therefore been connected to theories of prosodic phrasing.

Born out of the observation that syntactic structure (i.e. the parsing of utterances into phrases and clauses) and prosodic structure (e.g. the division of the speech stream in chunks which are separated by pauses) need not be isomorphic, the theory of *Prosodic Phonology* (Nespor & Vogel 1986, 2007) has been continuously expanded and refined over the last decades. In general, the derivational output of syntactic structure in this theory is interpreted by mapping rules into a phonological representation, which forms the basis for the application of phonological rules such as tone sandhi (Nespor & Vogel 1986; Selkirk 1984). Generally, this mapping can be based on the relationship between the syntactic constituents (Nespor & Vogel 1986), or the prosodic structure can be derived by mapping the edges of prosodic and syntactic constituents (Selkirk 1986).

Table 1.1 lists some of the commonly agreed levels of prosodic structure, which will be of relevance for the investigations in this thesis. They are presented alongside the syntactic structures which they are commonly associated with, but of course the details of the mapping diverge between different approaches. The specific correspondences will be discussed in the respective chapters of this thesis.

Table 1.1: Commonly assumed levels of prosodic structure.

Name	**Symbol**	**Syntactic structure**
Intonational phrase	*I*	Syntactic clause (CP)
Phonological phrase	φ	Maximal projection (incl. complements)
Prosodic word	ω	Lexical word

In order to empirically test the predictions made by the mapping algorithms, different aspects of the surface pronunciation have been taken into consideration. For example, under the assumption that prosodic domains are application domains for rule-based phonological changes on the segmental or tonal level, the non-application of these phenomena has been taken as indication that a prosodic domain boundary intervenes between the respective constituents. In this way, application or non-application of the phonological rule is taken as indicator for the presence or absence of the prosodic boundary (Baltazani 2006; Frascarelli

1999, 2000; Nespor & Vogel 1986; Selkirk 1986; Selkirk & Shen 1990; Truckenbrodt 1999).

Other approaches have assumed that the atomic units of intonational melodies (such as specific accents) are distributed over the sound string with reference to prosodic domains (Kahnemuyipour 2003; Kawahara & Shinya 2008; Pierrehumbert 1980; Pierrehumbert & Beckman 1988; Pierrehumbert & Hirschberg 1990). It is an open research question whether such prosodic domains, as demarcated by intonational events, are isomorphic with prosodic domains that are demarcated by e.g. final lengthening or the domains of application of phonological processes in all languages. Positive evidence in favor of this hypothesis has been put forward e.g. in Cao & Zheng 2006; Cho 2004; Cho & Keating 2009; Costa & Kula 2008; Dehé & Samek-Lodovici 2009; Dresher 1994; Frota 2000; Hayes & Lahiri 1991; Hellmuth 2007; Zheng et al. 2006, whereas Jun (1998) finds evidence for an opposing view.

1.1.3 Focus

One of the factors that influence tonal realization has to do with the discourse status of the constituent in question. This type of discourse-dependent modification has been subsumed under the cover term *information structure* (see e.g. Krifka 2007 for an overview). It covers phenomena such as *topic*, *focus*, and *givenness*. In this thesis, the two most important information-structural notions will be *focus* and *givenness*. Furthermore, the discussion will center on the phonetic or phonological reflexes of information structure, since these are the types of effect that are most relevant for the investigation of information structure in a tone language.

Generally, information-structural notions can be approached from two perspectives: (i) the semantic/pragmatic perspective, which takes the discourse specifications as primary, and (ii) the phonological/prosodic perspective, which starts from the phonetic or phonological reflexes of the information-structural categories. In this thesis, the second approach will be taken, and commonly agreed definitions of information-structural terms will be assumed in order to investigate the effects of information structure.

The most commonly investigated information-structural category, *focus*, has been defined according to different criteria. In this thesis, two types of focus are of interest. The first type, referred to as *presentational focus* (Gussenhoven 2007), *information focus* (Halliday 1967), or simply *focus* (Breen et al. 2010), is commonly considered to refer to the part of a sentence or utterance that corresponds to the wh-word in a related question. In contrast, the type of focus

that is induced by an explicit alternative in the discourse, which is corrected or contrasted in the utterance under question, is referred to as *identificational focus* (É. Kiss 1998), *contrastive focus* (Selkirk 2007b), or *corrective focus* (Gussenhoven 2007).

A notion that is somewhat orthogonal to these two types of focus is the notion of *givenness*. A given constituent is usually assumed to be activated in the discourse, either because it has been previously explicitly mentioned, or through entailment (Schwarzschild 1999). The two notions of *focus* and *givenness*, although situated on somewhat opposite ends of the spectrum, are therefore not mutually exclusive: a constituent can very well be simultaneously given and focused in a specific discourse situation (Büring 2006; Féry & Ishihara 2009).

With respect to focus, another common distinction is that between *broad*/*wide* and *narrow* focus (Breen et al. 2010; Eady et al. 1986; Ladd 1996, 2008). The distinction is often investigated with respect to wh-induced focus, whereby the focus domain corresponds to a single word or constituent under narrow focus, and to a phrase or the entire sentence under broad focus. It has been argued that the types of phonetic marking of focus in the same language may vary between broad and narrow focus (Hayes & Lahiri 1991), or between discourse-new and contrastive focus (Katz & Selkirk 2011).

Sometimes used with reference to the same domains that are referred to with *broad focus*, the notion of *all-new* (e.g. in Féry & Kügler 2008) or *out-of-the-blue* (e.g. in Katz & Selkirk 2011) refers to situations in which an entire sentence is focused as an answer to a general question such as "What happened?". Even though the broad focus sentences are not focus-free, they are sometimes included as a baseline or control condition in studies which compare different types of focus.

1.1.4 Tonal realization and prosodic structure

For dialects of Chinese, most studies that are concerned with the relation between prosodic structure and tonal realization have looked at categorical, tone-sandhi type tone changes. As described in section 1.1.2 above, the application patterns of the tone sandhi changes are taken to be indicative of prosodic structure, such that prosodic boundaries are assumed to intervene in contexts where tone sandhi changes fail to apply, and prosodic headedness is connected to the preservation of lexical tone values. Studies of this kind are Chen 1987; Lin 1994 for Xiamen, Brooke et al. 2009; Cheng 1987; Shen 1990a; Shih 1986, 1997; Zhang 1997 for Mandarin, Lin 2005 for Sixian Hakka and

Mandarin, Tsay & Myers 1996; Tsay et al. 1999 for Taiwanese, and Selkirk & Shen 1990 for Shanghai Chinese. Cross-dialectal comparisons such as Chen 1991, 2000; Duanmu 2005; Lee 2002; Yip 1999; Yue-Hashimoto 1987; Zhang 2007b are concerned with the tone sandhi application patterns in more than one Chinese dialect.

Outside of tone sandhi, tonal realization has also been shown to be influenced by intonational factors (Shen 1990a; Shih 2000; Yuan 2004), which in turn might be connected to prosodic structure. For example, in Yang & Wang 2002, it was shown that the realization of F_0 targets, such as F_0 minima, is connected to the prosodic structure of the sentences they appear in, such that there is a reset of F_0 minimum values which correlates in magnitude with the level of prosodic boundary that precedes the respective syllables. This indicates that prosodic structure might not just influence the selection of tonal targets (e.g. by blocking tone sandhi), but also have an effect on their implementation.

1.1.5 Tonal realization and focus

It has been instrumentally investigated for several dialects of Chinese how focus affects the realization of tones. Most studies have reported results on two acoustic parameters: F_0 and duration. In the studies mentioned below, the term *focus* is used to either refer to informational focus induced by a wh-word in a question-answer pair, or to contrastive focus. In the absence of studies showing a different effect of these two types of focus for any dialect of Chinese, the studies below will be grouped together concerning focus effects.

In all Chinese dialects that have been instrumentally investigated, focus has been found to induce lengthening of the focused syllable or constituent, and to some extent of neighboring syllables. Such studies include Chen & Gussenhoven 2008 for Standard Chinese, Pan et al. 2005 for Taiwan Mandarin, Chen et al. 2009; Xu et al. 2012 for Taiwan Min, Gu & Lee 2007a for Cantonese, Chen 2009 for Shanghai Chinese, and Wu & Xu 2010 for Hong Kong Cantonese. As shown in Chen 2006 for Standard Chinese and Chen 2009 for Shanghai Chinese, the lengthening effect of focus is to some extent dependent on the prosodically induced duration distribution within the focused word.

More relevant for tonal realization, in many dialects, focus has also been found to affect the implementation of tonal contours. In most studies, the focus effects on F_0 have been summarized as F_0 expansion, whereby e.g. a rising tone will start lower and end higher under focus than the same tone under broad focus. F_0 range expansion under focus has been reported for Mandarin Chinese (Wang & Xu 2006; Xu 1999), Cantonese (Gu & Lee 2007b), Shanghai Chinese (Chen

2009), and Taiwan Mandarin (Chen et al. 2009; Xu et al. 2012). The absence of a focus effect on tonal implementation has been reported in Wu & Xu 2010 for Hong Kong Cantonese and in Chen et al. 2009; Xu et al. 2012 for Taiwan Min.

In most, but not all of the dialects in which focus affects the implementation of tones, the on-focus F_0 expansion effect is accompanied by a F_0 compression effect on the post-focal stretch of the target sentence. Acoustically, this compression manifests itself mainly in lowering of the F_0 and intensity in the post-focal stretch. In a perception experiment with partially masked test sentences, as reported in Xu et al. 2004, it was established that post-focal compression may be utilized as an important secondary cue to the identification of the focus domain by the listeners.

Taking a closer look at the phonetic nature of both the on-focus F_0 expansion effect and post-focus compression, Chen & Gussenhoven (2008) and Chen (2010) argue that F_0 range expansion might not be the only way to account for the changes to tonal implementation that are caused by focus. Rather, the authors argue that tones are hyper-articulated under focus, and hypo-articulated in post-focal condition, as manifested by a reduced degree of distinctiveness of the tonal contour, and a greater influence of the preceding tones. Therefore, while the phonetic findings themselves are mostly uncontroversial, their exact interpretation in relation to models of tonal implementation is still subject to ongoing debate.

1.1.6 Prosodic structure and focus

In impressionistic accounts of focus effects in Chinese dialects, it has been observed that they resemble prosodic boundary effects, in the sense that they may also influence the application of tone sandhi (Brooke et al. 2009; Chen 2000; Selkirk & Shen 1990; Shih 1997). To account for the observed differences in tone sandhi application under focus, it has been proposed that focus can alter the prosodic structure of a sentence or utterance, and insert or remove boundaries so as to derive a prosodic structure that is coherent with the focus requirements.

Such proposals are reminiscent of focus realization theories that conceptualize focus effects as being mediated by prosodic structure (Büring 2010; Kabagema-Bilan et al. 2011; Selkirk 2007a; Truckenbrodt 1995, 1999; Zec & Inkelas 1990). In these accounts, focus is marked by local prominence, but this prominence is induced by alternations in the prosodic structure, rather than directly in the phonetic implementation. In that sense, a focused constituent would be phonologically strengthened, and the magnified implementation would

be a consequence of this kind of strengthening, rather than a purely phonetic reflex of an expanded F_0 range.

However, detailed phonetic investigations of boundary effects and focus effects have cast doubts on these types of indirect accounts of focus effects. For Mandarin, Chen (2004) has shown that the effect induced by an intonation phrase boundary onto the post-boundary syllable was confined to the onset of that syllable. In contrast, focus on the post-boundary syllable, which under the above hypothesis should have a similar effect, affected both the onset and rhyme of the syllable in question. For German and Japanese, Féry (2010) and Féry & Ishihara (2010) present experimental evidence that suggests that prosodic boundaries induce effects that are superficially similar, but different in detail, from the effects induced by focus.

In this sense, there seems to be accumulating evidence to suggest that the effect of focus is independent from the formation of prosodic domains or prosodic prominence assignment. Rather, it is proposed in Féry 2010; Féry & Ishihara 2010 that prosodic structure is derived exclusively on the basis of syntactic information. Focus on the other hand cannot modify the prosodic structure directly, but only affect the implementation of material within the prosodic structure.

1.2 Overview of this dissertation

1.2.1 Research objective

As has been outlined in the previous sections, the three areas of tonal realization, prosodic phrasing, and information structure are connected in numerous ways. Evidence from tonal implementation and tone sandhi, as well as observations from the effects of focus, have been used to motivate or disprove assumptions on how to derive prosodic structure from syntactic structure, and how to conceptualize their connection. At the same time, investigations of tonal realization under focus have helped to improve the understanding of the articulatory mechanisms behind tonal contour implementation.

This thesis attempts to broaden the empirical basis for the development and furthering of theories that are concerned with prosodic structure and with focus effects, as well as with their interactions. By investigating a tone language with a tonal phonology that has been described to depend both on prosodic and focus principles, this thesis aims to complement previous research on prosodic phrasing and focus, which has largely been conducted on the basis of findings from intonational and stress-accent languages. Lastly, by experimentally testing

some of the predictions made in the literature on prosodic phrasing and focus, the empirical accuracy of these theories is further investigated.

1.2.2 Research questions

The following research questions will be addressed in this thesis:

Tone sandhi and prosody (Chapter 3):

- Which factors determine the application of tone sandhi in disyllabic targets which are ambiguous between two prosodic structures?

Tone sandhi and focus (Chapters 3 and 4):

- Can the presence of contrastive focus in (only one of the syllables of) a disyllabic lexical compound block lexical tone sandhi?
- If not, how are the acoustic reflexes of focus distributed within the disyllabic lexical compound if only one of the syllables is focused, compared to focus on the entire disyllabic lexical word?

Tonal realization and prosodic structure (Chapter 5):

- Is the implementation of tonal contours affected by prosodic structure? If yes, which component of prosodic structure (prosodic boundaries/prosodic heads) is more important for the way tonal contours are implemented?

Tonal realization, prosodic structure, and focus (Chapter 5):

- Is the effect of prosodic structure on tonal implementation identical to the effect of focus?

Tone sandhi contour implementation and prosodic structure (Chapter 6):

- How are tonal contours implemented/scaled in sentences with different numbers of words per constituent? Is the scaling of the contours based on sentence or on constituent length?
- How does syntactic embedding affect the scaling of tonal contours? Which level of syntactic complexity is reflected in the tonal scaling?

Tonal realization and focus/givenness (Chapter 7):

- Is there a difference in tonal realization between constituents that are given, broadly focused, and narrowly focused?
- Do the speakers of Wenzhou use lexical means to mark referents in different discourse situations?

1.2.3 Experimental methods

The data presented in this thesis was obtained by means of experimental production tests. In the majority of experiments reported here, the speakers were asked to read out words and sentences that were presented to them in Chinese

characters, or to act out short question-answer dialogues, which attempted to set up certain communicative situations. In one of the reported experiments, speakers were presented with pictures and a context question, and asked to answer the context question by means of describing the picture. In all cases, the answers given by the speakers were recorded, and later measured and investigated statistically.

1.2.4 Outline

This thesis is composed of the following chapters: Chapter 2 introduces the language which is investigated in the remainder of the thesis, the dialect of Chinese spoken in Wenzhou. The chapter presents an overview of the relevant phonetic and phonological properties, particularly the phoneme and toneme inventory on syllables, and the word-domain tone sandhi processes. It also discusses some differences between the speech of the young speakers that were recorded for the experiments of this thesis and the previously published literature on segmental and tonal properties of Wenzhou Chinese.

Chapter 3 is concerned with the application domain for the disyllabic tone sandhi process of Wenzhou Chinese. As suggested by the literature, tone sandhi always applies in some types of constructions, but is variable in others. Chapter 3 specifically investigates two claims made in earlier research: (i) the application rate of tone sandhi in disyllabic verb-object constructions is related to the degree of lexicalization of these verb-object constructions, and (ii) the application of tone sandhi in disyllabic verb-object constructions can be influenced by the presence of focus. The chapter additionally tests the influence of communicative context on the application rate of tone sandhi. The main finding is that the communicative context exerts the most influence on the tone sandhi application behavior of the young speakers, while lexicalization only plays a subordinate role, and contrastive focus does not affect the tone sandhi application process.

Chapter 4 investigates the effect of contrastive focus on the tone sandhi contours, by varying the position of the focus domain with respect to the tone sandhi domain. In this way, it will be directly tested whether focus can “break up” the tone sandhi domain, either phonologically (by blocking tone sandhi) or phonetically (by inducing a stronger F_0 effect on the immediately focused syllable within the disyllabic structure). It is found that neither process occurs, and instead the phonetic reflexes of focus (expansion of the F_0 and duration of the tonal contour) are distributed evenly over the entire disyllabic domain. The

obtained results have important consequences for the conceptualization of the tone sandhi domain, and for theories of focus effects in Chinese.

Chapter 5 looks at the implementation of contour tones on monosyllabic words, and specifically at the degree to which the implementation of these contour tones is affected by adjacent tonal targets. Two factors were tested: (i) prosodic structure (prosodic boundaries and prosodic heads), and (ii) Focus. By directly comparing the influence of these two factors independently, the chapter also addresses the question whether the effects of these two factors are identical or cumulative. In this way, the chapter directly relates to the research debate outlined in section 1.1.6. As will be shown, prosodic headedness and focus independently induce a strengthening effect on the respective syllable, which maximizes the realization of the tonal contour and increases its independence from the coarticulatory influence of neighboring tones. The two effects are shown to be cumulative, which leads to an analysis that conceptualizes them as independent.

Chapter 6 examines the properties of sentential F_0 scaling in Wenzhou. The research question for this chapter has more often been asked for intonational languages, and concerns the pre-planning of sentence-level intonation. Particularly, the chapter inspects whether the height of F_0 peaks and valleys is related to sentence or constituent length, and how it reflects syntactic complexity in embedded clauses. In this way, it is analyzed how the syntactic structure is mapped onto the prosodic structure, and how the prosodic structure determines the tonal implementation on the phrase and sentence level. The findings show that constituent length, rather than sentence length, is the level of pre-planning of F_0 peak scaling in Wenzhou, and that minute details of syntactic embedding are reflected in the F_0 contour implementation.

Chapter 7 explores the lexical and phonetic correlates of narrow and broad focus and givenness. In two experiments, speakers are first tested on a picture elicitation task, which investigates the lexical properties of the structures they use to describe referents in different information-structural contexts. It is found that speakers systematically vary the definiteness and lexical choice of noun phrases in relation to the discourse status of the referents that are described with these noun phrases. A related experiment asks whether speakers, when they are more constrained in the lexical material in discourse, will systematically vary the acoustic cues to distinguish referents in broad focus from referents in narrow focus and given referents, which turns out to be indeed the case. Thus, the experiment determines that the notion of focus alone is not sufficient to

characterize the phonetic correlates of information-structure marking in Wenzhou Chinese.

Chapter 8 concludes with a summary of the experimental findings, puts them in cross-linguistic perspective, and gives an outlook for possible directions of future research.

Chapter 2

Wenzhou Chinese phonetics and phonology sketch

2.1 General introduction

2.1.1 Area and intelligibility of the Wenzhou dialect

Wenzhou Chinese is a dialect of Chinese which is spoken in the Wenzhou urban area, home to approximately 9.12 million people.[1] Of the varieties spoken in the area, the one spoken in the central district of Wenzhou (Lucheng) is commonly regarded as most representative. It is more or less mutually intelligible with the other varieties spoken in and around the city, although speakers from Lucheng can usually point out the idiosyncrasies of speakers from the outskirt districts such as Longwan and Ouhai.

The Wenzhou dialect, as the dialect of the Lucheng area in Wenzhou will be referred to hereafter, is considered to be representative of the Oujiang dialect subgroup (Fu & Fang 1985), which belongs to the larger dialect group of Wu dialects. The term "dialect" in this context is commonly used by Chinese linguists to describe the different sub-groupings of Sinitic languages within China. In Western linguistic terms, the subgroups would commonly be classified as individual languages, partially because mutual intelligibility between the "dialects" can be low (Tang & van Heuven 2007, 2009). This thesis will follow the nomenclature of "dialect" when talking about the subgroups of Chinese, and refer to the Chinese standard language as Standard Chinese.

2.1.2 Previous descriptions of the Wenzhou dialect

The Wenzhou dialect of Chinese has been included in recent cross-dialectal descriptions in Chinese, such as the description by Hou (1998) in the *现代汉语方言音库 [Modern Chinese Dialect Sound Database]*-Series, or the *当代吴语研究 [Studies in the Contemporary Wu Dialects]*-Volume by Qian (1992). These accounts describe the dialect from a comparative perspective, with Standard Chinese as a reference point. Recent descriptions which focus

[1] Data from the 2010 6th national population census of China, available at http://www.wzstats.gov.cn/2010rkpc/infoshow.asp?id=4336.

specifically on the Wenzhou dialect have been published by You (2002) and Zhengzhang (2008).

Modern Western linguistic accounts of Wenzhou Chinese are often concerned with subparts of the phonological system, such as the account of the tone sandhi phonology in Chen 2000, or the phonetic descriptions of the tone and tone sandhi patterns in Rose 2000, 2001, 2002, 2004. Both of these sources derive their observations from recordings made in the 1980s with middle-aged speakers, and are therefore likely to represent an older state of the dialect.

The current account of the Wenzhou dialect is based on the speech of young speakers from Lucheng Wenzhou. All speakers recorded for this phonology sketch, as well as the thesis at large, were between 18 and 29 years of age.[2] Due to ongoing language contact with Standard Chinese, which the young speakers learned from an early age, their way to speak Wenzhou differs quite substantially from the above-mentioned descriptions in many respects. For this reason, the upcoming chapter attempts to give an overview of the young speaker's segmental and tonal inventory, so as to lay a basis for the descriptions in the remainder of the thesis.

In the following, the sound inventory of Wenzhou will be illustrated with example words given in Chinese characters (Hanzi), as well as in broad IPA transcriptions. For the young speakers, the immediate association with Chinese characters is Standard Chinese, and when asked to pronounce an individual character in Wenzhou dialect, they are sometimes unsure of the "correct" pronunciation. In this description, it was attempted to select illustrative characters which would be recognized by most speakers to be associated with the given pronunciation. Most of the examples in the current sketch were picked from the words that were used as illustrations in previous published descriptions of Wenzhou.

Throughout this thesis, the lexical tones of Wenzhou will be represented by Chao numbers, following the numerical description of Wenzhou in You 2002. The tone trajectories and characteristics themselves will be discussed in section 2.4.

[2] More detailed information on the speakers can be found in the individual chapters. The current sketch is mainly based on recordings with two female speakers, who were both 25 years old. All illustrations of F0 and spectral values in this sketch were produced from recordings with either of the two speakers.

2.2 Phoneme inventory

2.2.1 Consonants

Table 2.1: Overview of consonantal phonemes and their place and manner of articulation in Wenzhou Chinese.

	Bilabial	Labiodental	Alveolar	Alveolo-palatal	Velar	Glottal
Plosive	p^h p b		t^h t d		k^h k g	(ʔ)
Affricate			ts^h ts dz	$tɕ^h$ tɕ dʑ		
Nasal	m		n	ɲ	ŋ	
Fricative		f v	s z	ɕ		h ɦ
Approximant		ʋ		j		
Lateral approximant			l			

2.2.1.1 Plosives

(1)	Citation forms	Hanzi	Translation
a.	*p^ha42*	派	'to send, dispatch'
b.	*pa42*	扮	'to dress up'
c.	*ba31*	排	'to arrange'
d.	*t^ha42*	太	'excessively, too'
e.	*ta42*	带	'belt, zone'
f.	*da31*	谈	'to talk, chat'
g.	*k^ha42*	快	'fast'
h.	*ka42*	界	'boundary'
i.	*gau24*	厚	'thick'
j.	*ʔa35*	矮	'short, low'

The plosive onsets of Wenzhou differentiate four different places of articulation: bilabial [p^h, p, b], alveolar [t^h, t, d], velar [k^h, k, g], and glottal [ʔ]. In terms of voicing, Wenzhou displays a three-way voicing distinction in plosives and affricates. This voicing distinction is correlated with tonal register in a way that the high register tones (H) only occur on syllables with voiceless and voiceless aspirated onsets, and the low register tones (L) only occur on voiced onset syllables. The minimal pair examples in (1) are limited by this tonal differentiation.

The exact phonetic nature of the three-member voicing contrast in Wenzhou and many other Wu dialects has attracted research interest for a long time. Instrumental studies going back to Chao 1928 have shown that none of the three "voicing" states is actually characterized by vocal fold vibration in the onset consonant when it is in word-initial position. Furthermore, it has been established for related dialects that in initial position, the difference of VOT between the so-called "voiced" and voiceless unaspirated onsets is negligible (Cao 1982; Shi 1983).

In the aerodynamic study by Cao & Maddieson (1992), it was shown that while there is no phonetic "voicing" contrast in Wenzhou onsets in initial position, the onsets that are traditionally described as "voiced" and "voiceless" do nonetheless differ in tenseness and aperture of the vocal folds. This causes a phonation difference at the release of the stops, and some breathiness at the onset of the following vowel, which aligns with the traditional terminology of "murmured" vowels (Duanmu 2007). According to Rose (2001), this difference conditions the so-called "depression" effect, which lowers the F_0 onset of the tonal contour following the "voiced" onset. Crucially, he argues that in Wenzhou this depression effect is independent of the register of the tone, and can also be found on tonal contours which have been changed to e.g. high falling in a tone sandhi environment.

Therefore, when talking about "voicing" in Wenzhou, several connected phenomena need to be kept apart. Phonetic voicing, as defined by a difference in VOT, is only found in word-medial position between the "voiced" and "voiceless" onset consonants and affricates, but perceptually neutralized in word-initial position. At the same time, the "depression" effect of the "voiced" onsets onto the beginning of the F_0 contour is present in both contexts. Finally, the limitation of syllables with "voiced" onsets to only carry low register tones is sometimes eliminated in tone sandhi context, but independent of the "depression

effect". For further discussion of the tone system of Wenzhou, the reader is referred to section 2.4 of this thesis.

In the examples in (1), the final example has been transcribed with a glottal stop [ʔ] onset. In Chinese linguistics, the existence of glottal stops in Chinese dialects is subject to debate for reasons of diachronic reconstruction (Zhang 2006). For this reason, syllables like those in (1j) have been alternatively transcribed as onsetless [a] (Zhengzhang 2008) or as having a zero onset [0a] (You 2002; You & Yang 1998) in Wenzhou, while other descriptions used the glottal stop onset [ʔa] (Qian 1992; Rose 2001).

For the young speakers described in this sketch, it could be observed that their realization of examples like (1j) can vary between realizations with a clear glottal contraction at the beginning, or with a more breathy transition into the vowel without glottal constriction (see Figure 2.1). This inter-speaker variation could be a reason for the different transcriptions of this sound. Since there is no phonemic alternation between onsetless syllables and those with a glottal stop onset, the glottal stop is set in parentheses in Table 2.1. For this sketch, in spite of the inter-speaker variation, the glottal stop will be consistently used to transcribe otherwise onsetless syllables.

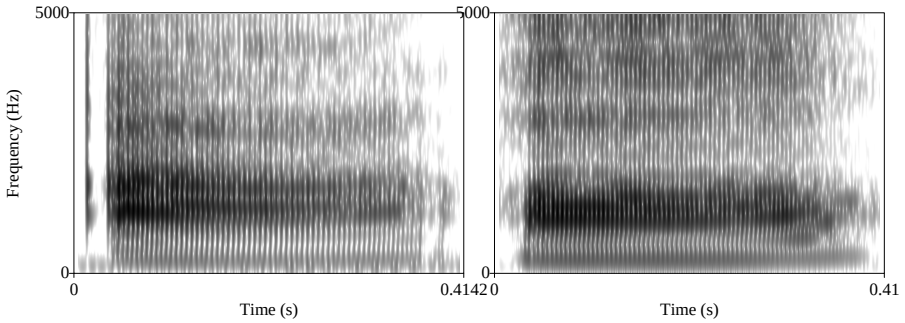


Figure 2.1: Realizations of example 1j (ʔa35) with (left) and without (right) glottal stop onset by two female speakers.

2.2.1.2 Affricates

(2)	Citation forms	Hanzi	Translation
a.	*ts^{h}a33*	差	'differ, wrong'
b.	*tsa42*	蘸	'to dip in liquid'

c.	*dza31*	惭	‘ashamed’
d.	*tɕʰa313*	雀	‘sparrow’
e.	*tɕa313*	脚	‘foot’
f.	*dzaŋ24*	近	‘near’

Affricates are produced with friction noise added to the release portion of an obstruent, and thus resemble both stops and fricatives to some extent. Wenzhou differentiates two places of articulation for affricates: alveolar [tsʰa, tsa, dza], and alveolo-palatal [tɕʰ, tɕ, dʑ]. With respect to onset voicing, the affricates of Wenzhou display the same three-member distinction that has been established for the plosives, and the same register division holds (low register tones with “voiced” affricate onset syllables, high register tones with voiceless and voiceless aspirated onset syllables).

The alveolo-palatal affricates [tʰɕ, tɕ, dʑ], like their fricative counterparts [ɕ, ʑ], induce a palatalization of the initial part of the following nucleus vowel if it is [–high], as demonstrated by the narrow transcriptions of the examples in (3): in (3a-c), in which the nucleus contains a [+high] vowel as first or only vowel ([ɵ], [u], and [i], respectively), no palatalized coarticulation can be found. In contrast, the palatal place of articulation of the onset affricates in (3d-h) is spread onto the beginning of the vowel articulation of the [–high] vowels.

(3)	Citation forms	Hanzi	Translation	Narrow transcription
a.	*tɕø33*	专	‘special’	[tɕø]
b.	*tɕuɔ33*	桩	‘stake, pile’	[tɕuɔ]
c.	*tɕiɛ33*	招	‘to attract’	[tɕiɛ]
d.	*tɕa313*	穿	‘to wear’	[tɕʲa]
e.	*tɕau42*	救	‘to rescue’	[tɕʲau]
f.	*tɕou33*	州	‘prefecture’	[tɕʲou]
g.	*tɕaŋ33*	金	‘gold’	[tɕʲaŋ]
h.	*tɕoŋ33*	中	‘middle; in’	[tɕʲoŋ]

In this sketch, the phenomenon will be regarded as regular onset-vowel coarticulation process for the palatal affricates and fricatives, and consequently not be indicated in the broad transcription. In traditional descriptions of other Chinese dialects, the palatal coarticulation between onsets and non-high vowels is sometimes transcribed with a triphthong starting with an [i] (e.g. [ɕiou], [ɕiaŋ] in the phonetic description of Standard Chinese in Lee & Zee 2003). However,

the duration of the palatal coarticulation in Wenzhou is very short, and only lasts less than 50 ms into the articulation of the vowel, as illustrated in Figure 2.2 For this reason, it will be treated as coarticulation here.

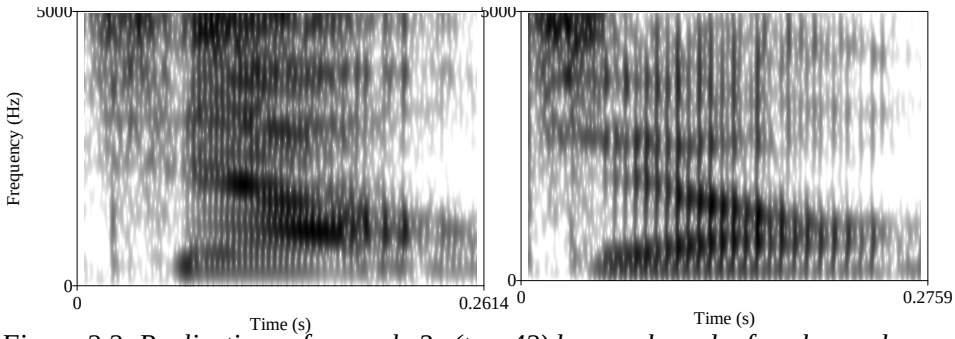


Figure 2.2: Realizations of example 3e (tɕau42) by a male and a female speaker.

2.2.1.3 Nasals

(4)	Citation forms	Hanzi	Translation
a.	*ma24*	买	'to buy'
b.	*na24*	乃	'knife'
c.	*ɲa24*	鸟	'bird'
d.	*ŋa24*	眼	'pupil of the eye'

The nasal series of Wenzhou contains four different places of articulation: bilabial [m], alveolar [n], palatal [ɲ], and velar [ŋ]. Traditionally, nasals are counted as sonorous onsets, and should therefore be expected to pattern with voiced obstruents and only co-occur with low register tones. However, according to the descriptions in Zhengzhang 2008 and the observations from the recordings with the young speakers, all four nasals can co-occur with both high and low register tones. An example for this is given in the minimal pair in (5k-l), and illustrated in Figure 2.3: the syllable [mi] with low and high level tone, as uttered by the same female speaker. The average F_0 in the two examples is ~220 Hz for the low level (5k), and ~245 Hz for the high level (5l).

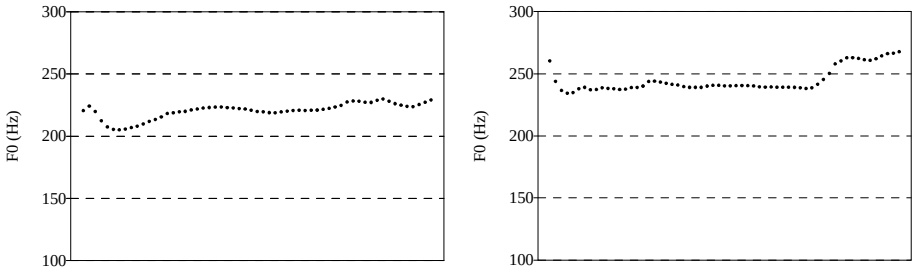


Figure 2.3: Realizations of example 5k (mi11, mean F_0 = 220 Hz) and 5l (mi33, mean F_0 = 245 Hz) by the same female speaker.

The velar nasal [ŋ] deserves special mentioning because of its phonotactic behavior: it can occur as onset consonant before monophthongs and diphthongs (5a-c), as coda consonant following the non-high vowels [a], [ɜ] and [o] (5d-f), and it can stand on its own as a syllabic nasal and carry different tones (5g-j).

(5)	Citation forms	Hanzi	Translation
a.	*ŋa31*	崖	‘cliff’
b.	*ŋu31*	牙	‘tooth’
c.	*ŋau31*	牛	‘ox’
d.	*maŋ31*	门	‘door’
e.	*mɜŋ31*	明	‘bright, clear’
f.	*moŋ11*	梦	‘dream’
g.	*ŋ11*	卧	‘to lie down’
h.	*ŋ24*	我	‘I’
i.	*ŋ24*	五	‘five’
j.	*ŋ31*	儿	‘child’
k.	*mi11*	面	‘face’
l.	*mi33*	眯	‘to narrow one's eyes’

In connected speech, such as phrases and sentences, words which only consist of a syllabified nasal can be merged with preceding (open) syllables and then resemble coda consonants (see also Pan 1991). Thereby, the actual syllable

boundaries can be obscured, which has consequences for the phonotactic analysis (see section 2.3.1 for details).

2.2.1.4 Fricatives

(6)	Citation forms	Hanzi	Translation
a.	*fa42*	泛	'flood'
b.	*va31*	凡	'ordinary'
c.	*sa42*	帅	'graceful'
d.	*za31*	柴	'firewood'
e.	*ɕa35*	晓	'dawn'
f.	*ha42*	喝	'to drink'
g.	*ɦa31*	鞋	'shoe'
h.	*ʋa33*	歪	'crooked'
i.	*va33*	弯	'curve, bend'

Fricatives in Wenzhou differentiate four different places of articulation: labiodental [f, v], alveolar [s, z], alveolo-palatal [ɕ], and glottal [h, ɦ]. Like for the plosives and affricates, there is a correlation between tonal register and onset voicing of the fricatives. As noted in Qian 1992, the bilabial fricatives are subject to ongoing language change in many Wu dialects, in the sense that younger speakers seem to use bilabial [f] and [v] more consistently, whereas older speakers vary between [v] and [β] and between [f] and [ɸ]. In the recordings with the young speakers here, the only noticeable variation occurred for [v], which was sometimes produced in a more approximant-like manner as [ʋ]. Since no phonemic alternation between an approximant [ʋ] and a fricative [v] has been reported in any previous description of Wenzhou, it is assumed here that the approximant is a free variation of the pronunciation of the phoneme [v]. The difference is illustrated with the two examples (6h) and (6i).

The alveolo-palatal fricative [ɕ] is the only fricative with no voiced counterpart in the fricative series of Wenzhou. It only co-occurs with high register tones, and can therefore be deemed "voiceless". Comparable to the alveolo-palatal affricates, [ɕ]-initial syllables induce palatalization of the beginning of the following vowel if it is [–high]. Narrow transcriptions of some examples in (7) illustrate this process, but in accordance with the discussion in the affricate section, the broad transcription in the remainder of this thesis will not reflect this distinction.

(7)	Citation forms	Hanzi	Translation	Narrow transcription
a.	*ɕø42*	汉	‘Chinese’	[ɕø]
b.	*ɕiɛ33*	稍	‘a little’	[ɕiɛ]
c.	*ɕuo33*	霜	‘frost’	[ɕuɔ]
d.	*ɕa35*	晓	‘dawn’	[ɕʲa]
e.	*ɕou33*	收	‘to gather’	[ɕʲou]

From a cross-linguistic point of view, the “voiced” glottal fricative [ɦ] is rare in many other language families of the world, but very common in the subgroup of the Wu dialects.

2.2.1.5 Approximants

(8)	Citation forms	Hanzi	Translation
a.	*ja212*	药	‘medicine’
d.	*la212*	兰	‘orchid’
c.	*la33*	拉	‘pull, drag’
d.	*la11*	烂	‘messy’

Like the nasals, the approximant onset consonants in Wenzhou can co-occur with both high and low register tones. Figure 2.4 illustrates the realizations of the examples in (8c) and (8d) by the same female speaker. The syllable [la] with high level tone (8c) has an average F_0 of ~214 Hz, while the low level example (8d) has an average F_0 of ~188 Hz.

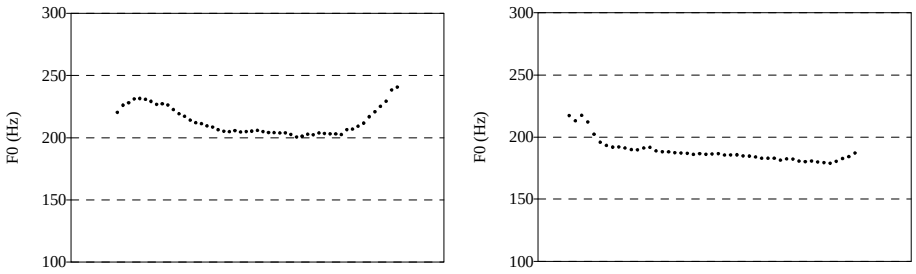


Figure 2.4: Realizations of example 8c (la33, mean F_0 = 214 Hz, left) and 8d (la11, mean F_0 = 188 Hz, right) by a female speaker.

2.2.2 Vowels

2.2.2.1 Monophthongs

Due to numerous co-occurrence constraints and some gaps in the syllable inventory, the velar plosives [g, k, k^h] are the only onsets that can co-occur with all monophthongs in Wenzhou. In (9), a series of (near) minimal pairs is given. The overview of possible Wenzhou syllables in Zhengzhang 2008 contains no example of [kɵ], which is why the nearest approximation [k^hɵ] is listed in (9).

(9)	Citation forms	Hanzi	Translation
a.	*ka35*	解	'to separate, explain'
b.	*kɜ35*	稿	'draft, sketch'
c.	*k^hɵ35*	砍	'to chop, hack'
d.	*ko35*	假	'false'
e.	*ku35*	果	'fruit'
f.	*kø35*	赶	'to catch up'
g.	*ki35*	几	'small table'

The quality of some of the vowels in Wenzhou is different from the corresponding vowels in Mandarin dialects. Unlike the [y]-vowel in the northern dialects for example, the vowel that is transcribed as [ø] here and as [y] in most descriptions of Wenzhou (Hou 1998; You 2002; You & Yang 1998; Zhengzhang 2008) is retracted, less rounded, and also lower in quality than its cognate [y] in Standard Chinese. Conversely, the sound that is transcribed as [ø] or [$ø^y$] in Hou 1998; You & Yang 1998; Zhengzhang 2008 is retracted for the young speakers recorded here and articulated in a central position, and will consequently be transcribed here as [ɵ]. For the vowel [ɜ], only one symbol will be used in this sketch, whereas the descriptions in Hou 1998; You 2002; You & Yang 1998; Zhengzhang 2008 use three different symbols ([ɜ], [ɛ], and either [e] or [ə]) to represent the phonetic value of the sound in different contexts.

It has been observed for a number of Wu dialects that the back high vowel [u] tends to have a tendency to be unrounded (Pan 1991), and that specifically in the Wenzhou phoneme inventory, "[u] has developed into /ʋ/ in the Wenzhou area, a vocalized version of /v/." (ibid: 260). For some of the Wenzhou speakers recorded here, there is indeed a tendency for less lip protrusion, so that the lips are more spread during the vowel production. This can co-occur with some (optional) lip trilling or breathing during the vowel

production, so that the vowel indeed assumes a more fricative-like timbre. Also, as in many other Wu dialects, the two phonemes [u] and [o] are closer together than in other languages, largely through a raising of [o]. The two phonemes remain distinguishable by native listeners, but may fall in the same phoneme category for listeners from Indo-European languages such as German or Dutch.

As in many other Chinese dialects, Wenzhou has a regular alternation for the sound that occurs after alveolar fricative and affricate onset consonants in complementary distribution with [i] in other contexts. This sound is traditionally transcribed with the <ɿ>-grapheme in Chinese linguistics, and is transcribed here as a syllabic fricative [z̩]. Phonetically, the tongue position of the onset fricatives is preserved throughout the pronunciation of the syllable, and merely the voicing is changed at the transition from onset to nucleus in the case of voiceless onset consonants (Ladefoged & Maddieson 1996). The transcriptions in (10) serve to illustrate the process, but this sketch remains agnostic in the debate on the "correct" representation of the "fricative vowels" of Chinese. For discussion, see e.g. Zhang 2006: 53-57 and Duanmu 2007.

(10)	Citation forms	Hanzi	Translation
a.	*ts^{h}z̩42*	处	'place'
b.	*tsz̩42*	至	'to, until'
c.	*dz̩31*	迟	'tardy'
d.	*sz̩42*	似	'similar'
e.	*zz̩31*	视	'to look'

In contrast to Mandarin, which allows a wide range of monophthongs and diphthongs before the nasal codas [n] and [ŋ], Wenzhou only allows the non-high vowels /a/, /ɜ/, and /o/ before the velar nasal [ŋ], which is also the only possible coda consonant in the dialect. In the combination with this coda consonant, the realization of /ɜ/ and /o/ is centralized for all three vowels, compared to the monophthong realization in open syllables. This is reminiscent of the difference between vowels occurring in open and closed syllables in Shanghai Chinese, where the vowels exhibit a reduced acoustic vowel space in closed syllables (Chen 2008). A narrow transcription of the examples in (11) would transcribe them as [mɐŋ], [məŋ], and [moŋ], with the latter being lower and more front than the [o] in open syllables.

(11)	Citation forms	Hanzi	Translation
a.	*maŋ11*	闷	'stuffy'
b.	*mɜŋ11*	命	'life'
c.	*moŋ11*	梦	'dream'

2.2.2.2 Diphthongs

(12)	Citation forms	Hanzi	Translation
a.	*t^hai33*	推	'to push'
b.	*t^hɛi33*	梯	'ladder'
c.	*t^hau33*	偷	'to steal'
d.	*t^hou33*	拖	'to pull, drag'
e.	*t^hiɛ33*	挑	'to choose'
f.	*t^huɔ33*	汤	'soup'

As can be seen in (12), when counting [ɵ^{y}] as an alternative realization of the monophthong [ɵ], Wenzhou differentiates six diphthongs. Most Chinese grammars further differentiate combinations of the diphthongs with an initial [i] or [ø], thus deriving additional diphthongs with onglides or triphthongs. In the present investigation, they will be regarded as regular allophones, in accordance with the above analysis of the variants of [ɕ] inducing palatalization in [–high] vowels (cf. section 2.2.1.4, examples in (7)).

2.3 Phonotactics

2.3.1 Well-formed syllable structure

Since the velar nasal coda in Wenzhou can only co-occur with the monophthong vowels [a], [ɛ] or [ɔ], the maximal syllable in Wenzhou is *CV*(*X*), with *X* being either a second vowel (in diphthongs) or a velar nasal. Classifying Wenzhou as having an obligatory onset presupposes the analysis of the 0-onset as glottal stop (see section 2.2.1.1 for discussion), and the analysis of affricates as single *C*. Furthermore, as discussed in section 2.2.1.3, the velar nasal can be syllabic, and obligatorily be combined with *CVV* syllables. Unlike the [ɚ]suffix in Beijing Mandarin and other Mandarin dialects though (Duanmu 2007), the [ŋ] in Wenzhou does not merge with its host syllables, but is merely added to their

existing structure. This may create superheavy *CVVN* syllables, as in example (13).

(13)	Citation forms	Hanzi	Translation
a.	*muɔ33*	猫	'cat'
b.	*muɔŋ31*	猫儿	'cat'

2.3.2 Attested combinations of consonants and vowels

In line with other Chinese dialects, Wenzhou displays certain co-occurrence restrictions when it comes to the combination of the phonemes of the language into valid syllables. Some of these have already been covered, such as the allophony on [+high, +front] vowels following alveolar fricatives and aspirates (section 2.2.2.1), and the restriction of the coda consonant [ŋ] to only co-occur with allophones of three of the monophthong vowels in single syllables (sections 2.2.2.1 and 2.3.1). Outside of these, the only consistent co-occurrence restriction that could be found across all possible onset-rime combinations is a restriction for alveolo-palatal onsets to occur before rimes that start with mid vowels ([ɜ], [ɵ], [ɛi], [ɜŋ].)

However, this does not mean that all other possible onset-rime combinations indeed occur in Wenzhou. Appendix 2.2 gives an overview of attested syllables, and illustrates some mismatches between the syllables listed in the most recent comprehensive description of Wenzhou (Zhengzhang 2008) and the recordings with the young speakers for this thesis. These mismatches occurred in both directions, with some speakers realizing characters with syllables that are not listed in Zhengzhang 2008, and with speakers reporting in elicitation sessions that they could not think of a word in Wenzhou that would be pronounced with a particular onset-rime combination that is listed as possible syllable in Zhengzhang 2008. It is unclear whether these mismatches constitute mispronunciations on the part of the speakers, or whether they represent ongoing language change. For cases where no word with a particular onset-rime combination could be elicited, but where the speaker did not deny the existence of the particular combination, the syllable is listed in appendix 2.2 as existent.

2.4 Tone inventory in monosyllables

(14)	Syllable	Tone type[3]	MC class	Hanzi	Translation
a.	*ta33*	high level	Ia	单	'alone, single'
b.	*ta35*	high rising	IIa	胆	'guts'
c.	*ta42*	high falling	IIIa	带	'belt, zone'
d.	*ta313*	high dipping	IVa	搭	'to put up, hang over'
e.	*da11*	low level	IIIb	潭	'pond'
f.	*da24*	low rising	IIb	淡	'thin, light'
g.	*da31*	low falling	Ib	谈	'to chat'
h.	*da212*	low dipping	IVb	踏	'to step on sth.'

The examples in (14) above illustrate the eight citation tones of Wenzhou. As mentioned in the description of plosive onsets in Wenzhou, the eight tones of Wenzhou are commonly split into two registers, each containing four tones. The register division aligns to some extent with the "voicing" properties of the onset consonant, but some onsets, such as nasals and approximants, can occur with both high and low register tones. It has been proposed that the register division arose from a spread of the voicing properties of the onset onto the vowel, and that the pitch differentiation occurred as a secondary process via tonogenesis (Duanmu 1992). However, the connection between the register of a tone and its height is non-automatic, and in some Chinese dialects, "high" register tones may in fact be phonetically lower than "low" register tones (Chao 1967, see also Hayward et al. 2003).

For the Wenzhou "high" and "low" register tones, a difference in tone height can be confirmed at least for the level and falling tones, so the traditional terminology will be kept for ease of reference. However, as the description of the acoustic properties of the tones shows, it is overly simplistic to think of the eight tones as four contours being realized in two different areas of the speakers'

[3] In order to facilitate reference to the eight tones in Wenzhou, they are described using a combination of the register (high/ low) and the tone type (level, rising, falling, dipping), which are commonly used translations of the Middle Chinese tone classes (e.g. in Rose 2001 and Chen 2000, the latter of who substitutes "even" for "level"). However, the terms should not be taken to be descriptive of the tonal trajectories: As already noted in Rose 2008, the "dipping" tones of the young speakers largely lack the final rise, and realize the "dipping" tones with a falling contour instead.

pitch ranges, as the terminology might suggest. For this reason, the properties of the tones will be described in as much detail as possible.

In traditional Chinese descriptions, tones are often transcribed with the help of the number system after Chao (1948, see Duanmu 2007 for critical discussion). However, the exact Chao number values that are used to describe the tones in a dialect may vary between transcribers, for example whether e.g. a falling tone is described as 53 or 42, and whether tone duration is reflected in the number system (as for example in the description in Rose 2000). Table 2.2 lists some Chao number descriptions of Wenzhou tones that have been published in the recent literature, before describing in detail the tonal realization observed in the young speakers who were recorded for this phonology sketch.

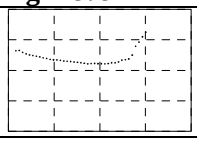
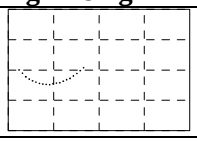
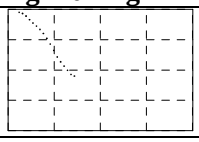
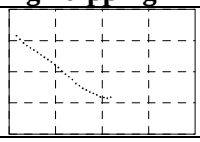
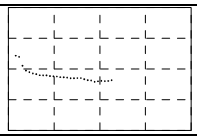
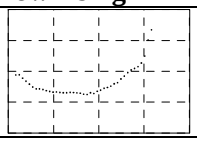
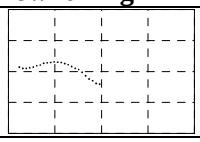
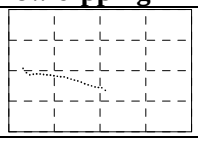
Table 2.2: Numeric descriptions of Wenzhou citation tones in the recent literature.

	High level	**High rising**	**High falling**	**High dipping**	**Low level**	**Low rising**	**Low falling**	**Low dipping**
Chen 2000	M (33)	MH (35)	HM (53)	Lq (313)	L (11)	MH (35)	ML (31)	Lq (212)
Nakajima 1983	44	45	42	323	31	23	242	212
Norman 1988	44	45	42	23	11	24	31	12
Hou 1998	33	35	42	213	22	35	31	213
Qian 1992	44	35	52	423	22	24	31	323
Rose 2000	33	34	52	3312	222	114	331	2212
You 2002; You & Yang 1998	33	35	42	313	11	24	31	212
Zhengzhang 1964b	44	45	42	323	22	34	31	212
Zhengzhang 2008	33	45	42	313	11	34	31	212

In order to illustrate the tonal contours of the young Wenzhou speakers, Table 2.3 depicts the F_0 trajectories of the examples in (15) as uttered by a female speaker. In order to allow for a comparison not only of the F_0 contour and height, but also of the durational properties of the tones, the pitch tracks are depicted

initially in a window of 400 ms length, with the left edge of window being aligned with the plosive burst of the onset consonant.

Table 2.3: Representative F_0 tracks of the eight lexical tones of Wenzhou on alveolar plosive onset syllables (examples 15a-h), recorded with a female speaker. F_0 range 100-300 Hz, intermediate lines = 150, 200, 250 Hz and 0.1, 0.2, and 0.3 seconds.

High level	**High rising**	**High falling**	**High dipping**
			
Low level	**Low rising**	**Low falling**	**Low dipping**
			

Visual and auditory inspection of the lexical tones of Wenzhou yield the following observations.

(15) **Observations on lexical tones in young Wenzhou speakers**

- The level tones are both situated in the middle of the speakers' pitch ranges, with notably the "high" level tone having a rather "mid" quality. Especially for the male speakers, the resulting shortage of distinctiveness is often compensated by lengthening of the low level tone. Both tones display a declining tendency, but for the female speakers, the high level tone may be accompanied by a short rising offset.
- Both rising tones are characterized by falling-level trajectories through the first half of the tonal duration. The "high" and "low" rising tones are often realized on similar F_0 height, which led e.g. Rose (2008) to claim that in the young speakers, the two tones have merged. However, in the young speakers recorded here, they are often distinguished by lengthening of the low register tone.
- For the falling tones, the distinction between "high" and "low" register is most clearly visible in the beginning of the tonal contour, which is

characterized by a short rising portion before falling to a comparable tone height as the high falling tone. Both tones are rather short of duration, most notably the high falling tone.

- For the dipping tones, the most striking observation is the lack of the final rise of the "dipping" contour. As the falling tones, the two dipping tones are distinguished mainly by the onset height of the fall, which is lower for the low-register tones.

As can be seen from the above description, several factors have to be taken into account when characterizing the tone contours of Wenzhou. Besides the actual tonal directions (which do not suffice any more to distinguish the falling from the dipping tones), the length of the tones and the presence/absence of the "depressor" effect at the onset of the tone are important for the distinction between "high" and "low" register tones.

2.5 Tone sandhi

2.5.1 Terminology

The term "tone sandhi" has been used in the literature with reference to two different phenomena: tonal coarticulation and rule-based tonal change (Chen 2000). The former phenomena include processes as described for example in Xu 1997 for Standard Chinese, involving assimilatory/dissimilatory and transitory processes that take place between two adjacent tones when they are produced in the context of one another. On the other hand, the term "tone sandhi" has also been used to refer to tone change processes in which at least one of the involved tones changes its phonological shape in a categorical manner. A famous example of such a change is the so-called "third tone sandhi" of Beijing Mandarin, in which, according to the common analysis, a third tone followed by another third tone within a specific prosodic domain is realized as a second tone (for discussions pro/contra full phonological neutralization see e.g. Chen & Yuan 2007; Peng 2000; Xu 1997; Zhang & Lai 2006, 2010).

These two uses of the term "tone sandhi" have been conflated on the basis of the argument that the two processes are not fundamentally different, and that they should be treated as gradual variants of each other (Chen 2000). However, as will become clear in the remainder of this thesis, for Wenzhou, it clearly makes more sense to keep the two terms apart. Therefore, the term "tone sandhi" will be used in this thesis to refer exclusively to such tone change phenomena in which, as stated above, at least one tone changes its phonological

shape in a categorical and rule-based manner. This implies that (i) the tonal change cannot be entirely explained as articulatorily conditioned modification of the shape of either of the lexical tones, and (ii) that the process is stable over different segmental TBUs, as long as they share similar morphosyntactic or prosodic conditioning environment.

While "tone sandhi" and "tonal coarticulation" are thus differentiated in terminology, both processes undoubtedly may serve the same function, namely simplification of articulation. Often, phonological accounts of tone sandhi processes strive to lend them motivation by referencing how they promote ease of articulation, or contribute to the reduction of markedness. Therefore, the above differentiation should be regarded as distinction of definition, but it does not exclude the possibility of functional overlap.

2.5.2 Domain of application of tone sandhi in Wenzhou Chinese

Commonly, the term "tone sandhi" with respect to Wenzhou Chinese describes processes targeting lexical words which consist of more than one syllable. The most common tone sandhi cited in the literature involves disyllabic lexical compound nouns, verbs, and adjectives, and it has also been argued to apply to disyllabic modifier constructions (Chen 2000). In the recordings of the young speakers, this pattern could be confirmed, as the speakers realized disyllabic lexical compound nouns, verbs, and adjectives with regular disyllabic contours, as described in the following section. The same was true for some adjective-noun-collocations. In contrast, in verb-object-collocations with two monosyllabic lexical items, the most prevalent pattern was the adoption of a phrasal prosody that retained the tonal value of the strong prosodic position (non-head) and reduced the tonal value of the prosodically non-strong position to low level. For an in-depth analysis of the domain of application of disyllabic tone sandhi, the reader is referred to Chapter 3 of this dissertation.

2.5.3 Patterns of tone sandhi contours in Wenzhou Chinese

A description of the patterns of (disyllabic) tone sandhi in Wenzhou has represented a challenge for phonologists, since the observed patterns seem to be "[d]efying explanation" (Rose 2004, citing Chen 2000). What is meant by that is that the connection between the citation forms and the sandhi contours in connected structures do not seem to be easily explicable through standard phonological processes such as spreading/shifting of tones, substitution, or neutralization (Chen 2000). Even if some of the patterns can be explained in a

rule-based manner, no unifying explanation has been proposed which could account for all observed tone sandhi contours (see also Yue-Hashimoto 1987).

For the current description, the matter is further complicated by the observation that the tone sandhi contours that were recorded with the young speakers do not fully match the tone sandhi contours that have been described even in the most recent literature (e.g. Chen 2000; Rose 2000, 2001, 2002, 2004; You 2002). This is caused partially by mismatches in the realization of the citation forms of the individual syllables, some of which are realized by the young speakers in a way that differs from previous descriptions on a segmental and/or on a tonal level. On the other hand, the observed tone sandhi contours are also clearly different even for those disyllabic examples where the realizations of the citation tones match published descriptions. In appendix 2.1, an overview is given of differences between some of the published descriptions and the tone sandhi contours recorded for this thesis.

Table 2.4 gives an overview of the tone sandhi contours that were observed in the young speakers, based on a broad transcription in Chao numbers.

Table 2.4: Numerical overview of tone sandhi contours observed in the speech of the young Wenzhou speakers.

<table>
<tr><th rowspan="2">First σ</th><th colspan="8">Second σ</th></tr>
<tr><th>33</th><th>31</th><th>35</th><th>24</th><th>42</th><th>11</th><th>313</th><th>212</th></tr>
<tr><td>33</td><td colspan="3">42.31</td><td>44.22</td><td rowspan="2">22.44</td><td rowspan="2">22.35</td><td rowspan="8" colspan="2">35.42</td></tr>
<tr><td>31</td><td colspan="2">31.22</td><td colspan="2">35.22</td></tr>
<tr><td>35</td><td>44.22</td><td>42.31</td><td colspan="2">44.22</td><td>42.31</td><td>44.22</td></tr>
<tr><td>24</td><td colspan="2">31.22</td><td colspan="2">35.22</td><td colspan="2">31.22</td></tr>
<tr><td>42</td><td>44.22</td><td>42.31</td><td colspan="2">44.22</td><td>42.31</td><td>44.22</td></tr>
<tr><td>11</td><td>35.22</td><td>31.22</td><td colspan="2">35.22</td><td>31.22</td><td>35.22</td></tr>
<tr><td>313</td><td rowspan="2">22.33</td><td rowspan="2">42.31</td><td rowspan="2" colspan="2">22.35</td><td rowspan="2">22.42</td><td rowspan="2">42.31</td></tr>
<tr><td>212</td></tr>
</table>

As can be seen from Table 2.4, even though the tone sandhi contours are different from the descriptions in the literature, there is still some neutralization in the patterns, in that specific tonal combinations result in tone sandhi contours that are acoustically comparable to the tone sandhi contours of other tonal

combinations. In (16), a verbal description of these tone sandhi contours and the tonal combinations they result from is listed.[4] The individual patterns are described in greater detail in the corresponding sections below.

(16) **Types of tone sandhi contours**

a. Rise – fall contour
→ x+dipping

b. Fall – low level contour
→ x+31, rising+42, 42/11+42, 33+33, 31+33, 33+35, 24+33, 24+11, dipping+11

c. Rise – low level contour
→ 31+rising, 24+rising, 11+rising, 11+33

d. Low level – rise contour
→ dipping+rising, 33+11, 31+11

e. High level – low level contour
→ 33+24, 35+rising, 42+rising, 35+33, 35+11, 42+33, 42+11

f. Low level – high level contour
→ 33+42, 31+42, dipping+33

g. Low level – fall contour
→ dipping+42

2.5.3.1 Rise – fall contour (16a)

As can be seen in Table 2.4, the category of dipping tones (212, 313) is the only tone category that displays regularity when appearing on the second syllable in a tone sandhi context. Combined with any other tone on the first syllable, the dipping tones result in a rise-fall tone sandhi contour (= pattern (16a, Figure 2.5).

[4] To differentiate the combinations of citation tones on the individual syllables from the disyllabic tone sandhi contours, the former will be indicated with a plus sign combining the citation tone values of the individual syllables, while the latter will be described with a dot between the tone sandhi contour values of the two syllables.

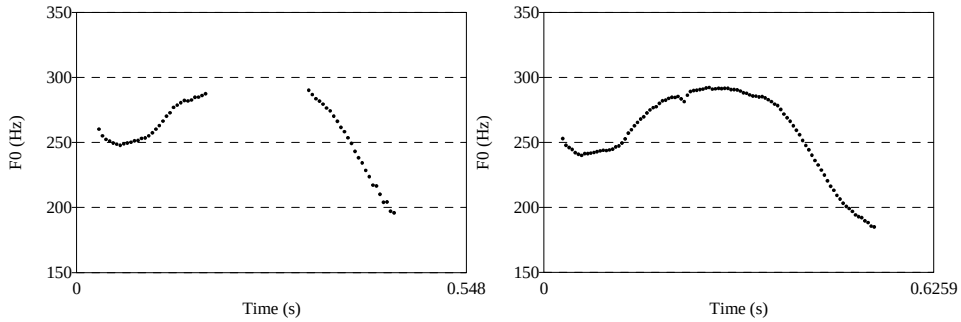


Figure 2.5: Representative pitch tracks of the rise-fall tone sandhi contour with a high register dipping tone (left) or a low register dipping tone (right) on the second syllable. Citation tones left = 35+313, right = 35+212.

In the tone sandhi contour resulting from x+313 combinations, the F_0 contour of the first syllable is characterized by a rise that starts about mid-level of the pitch range of the speaker and remains level for a short period of time, before rising to a high pitch point. The second syllable starts approximately at that point, falls sharply and straight throughout the second syllable, and ends at a pitch point that lies lower than the pitch onset point of the rise on the first syllable. The tone sandhi contours that result from x+212 combinations are characterized by a continuous transition from the rising to the falling part of the contour. Other than that, the pitch trajectory of the x+212 combinations mirrors that of the x+313 combinations, in that the pitch onset of the rise on the first syllable is higher than the pitch offset of the fall on the second syllable.

2.5.3.2 Fall – low level contour (16b)

A second very common tone sandhi contour in the speech of the young Wenzhou speakers involves a falling tone on the first syllable and a low level or declining tone on the second syllable (= pattern 16b, Figure 2.6).

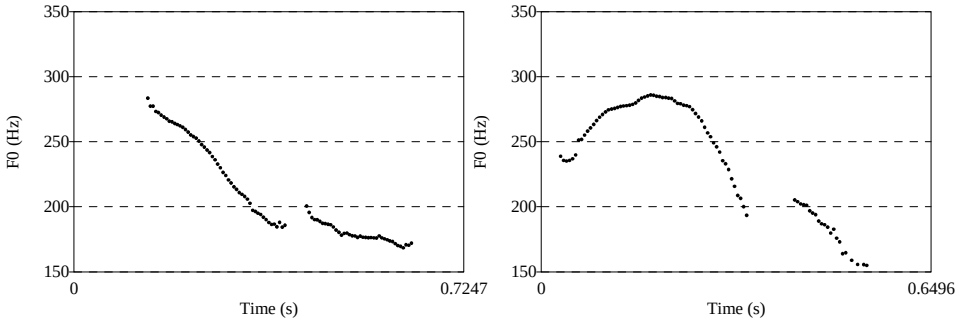


Figure 2.6: Representative pitch tracks of the fall-low level tone sandhi contour with a high register tone (left) or a low register tone (right) on the first syllable. Citation tones left = 33+31, right = 24+31.

Depending on the register on the first syllable, the onset of the fall on the first syllable will be either continuously falling (high register) or rising-falling (low register). In that sense, the shape of the falling tone contour on the first syllable is very similar to the lexically high and low falling tones, which are characterized by a similar difference in tonal direction (cf. Table 2.3 in section 2.4.1). This tone sandhi pattern occurs on all combinations with a low falling tone on the second syllable, and on four combinations with a high falling tone on the second syllable. Outside of these combinations of citation tones, however, the pattern also results in some combinations in which the citation tone on neither syllable is falling. It is therefore difficult to relate this tone sandhi contour to the citation tones on the individual syllables in a transparent way.

2.5.3.3 Rise – low level contour (16c)

Another frequent tone sandhi contour consists of a rising tone on the first syllable, followed by a low level or declining tone on the second syllable (= pattern 16c, Figure 2.7).

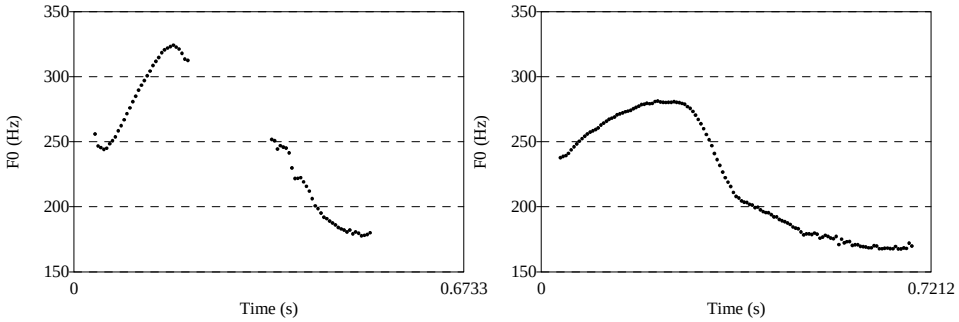


Figure 2.7: Representative pitch tracks of the rise-low level tone sandhi contour with a high register tone (left) or a low register tone (right) on the second syllable. Citation tones left = 11+35, right = 31+24.

As with the previous pattern, it can result from a combination of a syllable with a rising citation tone in second position, but since this tone sandhi pattern is also found on the combination 11+33, again no straightforward generalization can be made. This tone sandhi pattern shows some resemblance with the depressed falling-low level pattern described above. However, as can be seen from a comparison between Figures 2.6 and 2.7, the turning point in the connected falling-low level pattern is earlier than that in the connected rising-low level pattern. This suggests that the two patterns are not variants of each other, but indeed consist of different tonal contours.

2.5.3.4 Low level – rise contour (16d)

The inverse pattern, with a rise on the second instead of the first syllable, is found on the combination of dipping tones on the first syllable with rising tones on the second syllable, as well as a couple of other citation tone combinations (= pattern 16d, Figure 2.8).

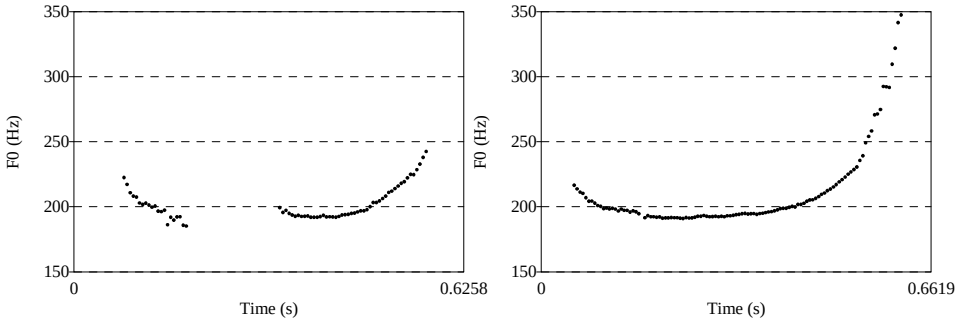


Figure 2.8: Representative pitch tracks of the low level-rise tone sandhi contour with a high register tone (left) or a low register tone (right) on the second syllable. Citation tones left = 313+35, right = 313+24.

2.5.3.5 High level – low level contour (16e)

Another tone sandhi contour that can be found across different combinations of citation tones is the combination of a high and a low level tone, mostly as result of a combination of two contour citation tones (= pattern 16e, Figure 2.9).

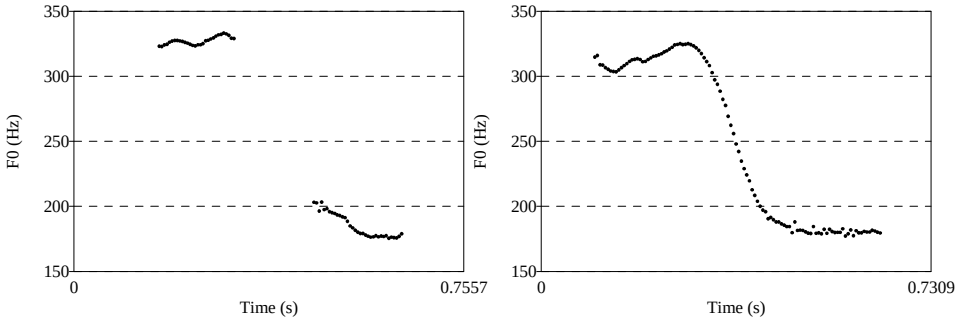


Figure 2.9: Representative pitch tracks of the high level-low level tone sandhi contour with a high register tone (left) or a low register tone (right) on the second syllable. Citation tones left = 35+35, right = 33+24.

One of the speakers of this sketch classified the high level tone of this tone sandhi contour, when played in isolation, as a high falling tone. This is not surprising, since it possesses two characteristic aspects of the latter, namely the shortness and the high F_0 level. In contrast, as discussed in section 2.4.1, the so-

called "high" level lexical tone of Wenzhou is situated in the middle of the speakers' pitch ranges, and lower than the high level tone in the sandhi contour discussed here.

2.5.3.6 Low level – high level contour (16f)

The inverse pattern to the just mentioned tone sandhi contour, namely a low combined with a high level tone, is only found in two combinations of citation tones, namely the high level or low falling tone on the first syllable plus the high falling tone (42) on the second syllable (= pattern 16f, Figure 2.10).

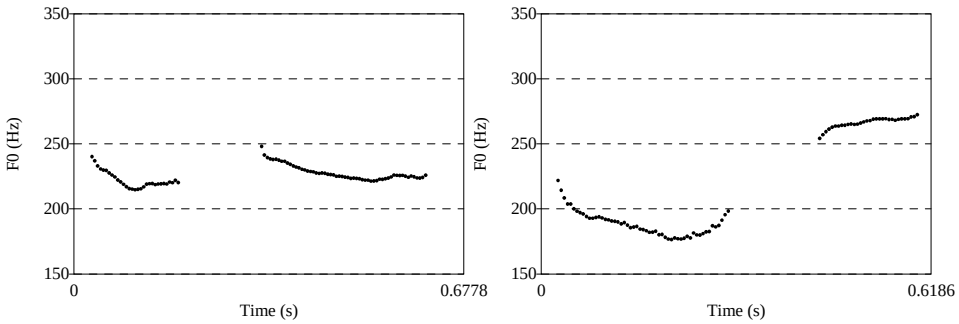


Figure 2.10: Representative pitch tracks of the low level-high level tone sandhi contour with a high register tone (left) or a low register tone (right) on the first syllable. Citation tones left = 313+33, right = 31+42.

As in the previous tone sandhi contour, the pitch trajectory of both syllables is level, but the high level tone on the second syllable does not display the extra-high F_0 target that it did in the previous pattern. Rather, it is situated in the mid range of the speaker's F_0 range, similar to the lexical "high" level tone.

Within the tone sandhi pattern (16f), two types of combinations of citation tones pattern together: 33/31+42, and dipping+33. While they are grouped into the same contour type in this analysis, an important distinction can still be made between these two types of citation tone combinations: the resulting length distribution in the tone sandhi contours is different. This can also be seen in Figure 2.8: the combinations of dipping+33 result in a short-long tone sandhi contour, while the combinations 33/31+42 result in a long-short tone sandhi contour. A possible explanation for this can be found in the underlyingly short duration of the high falling tone (42), which might be preserved in sandhi

contours even when the originally falling contour is replaced by a different tonal movement.

2.5.3.7 Low level – fall contour (16g)

The last tone sandhi contour results from the combination of a dipping tone on the first syllable and the high falling tone on the second syllable, and turns out as low level-high falling (= pattern 16g, Figure 2.11).

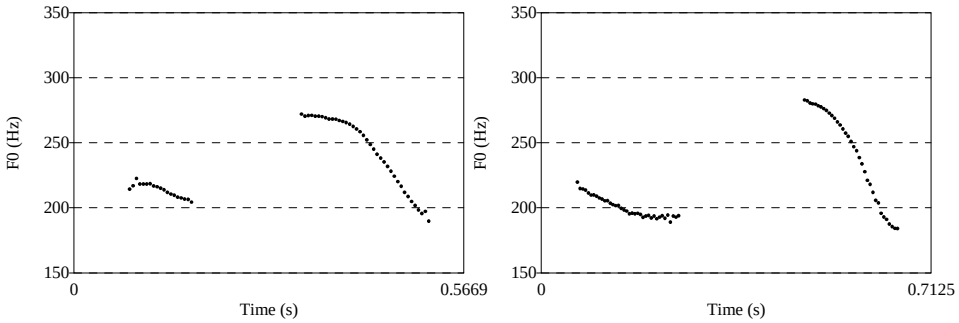


Figure 2.11: Representative pitch tracks of the low level-falling tone sandhi contour with a high register tone (left) or a low register tone (right) on the first syllable. Citation tones left = 313+42, right = 212+42.

2.5.4 Generalizations on the tone sandhi contours in Wenzhou Chinese

In the foregoing section, the tone sandhi patterns were grouped and described according to the similarities in the resulting tone sandhi contours. Abstracting away from the individual contours, this section will focus on the relationship between the citation tones on the individual syllables and the tone sandhi contours in different contexts, and discuss the occurring transparent connections.

When looking only at the combinations with dipping tones on the first syllable, it becomes apparent that the second syllables following the dipping tones mostly preserve their citation forms, while the first syllable often takes on a low level trajectory. This has led to generalizations in the literature in which it is claimed that in combinations of a dipping tone with any other tone, both syllables preserve their citation tone values, and the dipping tone trajectory becomes flattened and shortened. However, with the younger speakers, this generalization only holds true when excluding the patterns in which two dipping

tones combine. These latter patterns are realized with a rise-fall contour (= pattern 16a).

Another research question connected to the above mentioned observation concerns the distribution of "prominence" between the syllables of a word. A common assumption is that in disyllabic words, one syllable is "weak" whereas the other is "strong", and that this imbalance in prominence conditions or restricts the application of tone sandhi. Wenzhou has typically been regarded as a "right-prominent" language (Chen 2000; Yip 1999). This would entail the expectation that in the tone sandhi patterns, the rightmost syllables should generally be longer and carry more complex tonal contours.

As can be seen in the descriptions of the tone sandhi contours above, this generalization does not describe the findings for the young speakers accurately. While it is true that some tone sandhi patterns display an uneven distribution of duration that is independent of the segments of the syllables, these uneven distributions cannot be generalized across the entire data set. Rather, they seem to be connected to certain types of tones. For example, a dipping tone on the first syllable, combined with anything else than a dipping tone on the second syllable, will come out as a short low level tone in the tone sandhi contour. However, the same tone combined with another dipping tone will surface in an tone sandhi contour which is balanced for duration. Similarly, syllables with lexically falling tones can come out in tone sandhi contours as short when appearing either on the first or second syllable, but this is not a necessary process.

A couple of generalizations can be observed in the tone sandhi patterns of the young speakers, which are summarized in (17) and numerically in Table 2.5. The descriptions are verbal and no Chao numbers are given, so that tones with the same shape and different registers can be grouped together more easily.

(17) **Generalizations about tone sandhi patterns in young Wenzhou speakers**

a. Any tone combined with a dipping tone on the second syllable will result in a rise-fall contour.

b. A dipping tone on the first syllable combined with any other tone except another dipping tone on the second syllable will result in a pattern in which the first syllable is short and low level, and the second syllable retains its citation tone. In the case of the low falling and low level tone on the second syllable, the first syllable will become high falling, which results in a continuously falling contour over the two syllables.

c. A combination of any tone on the first syllable and a low falling tone on the second syllable will also always result in a continuously falling contour over the two syllables.

d. A high falling tone on the second syllable will result in a continuously falling contour over the two syllables when combined with any tone except the dipping tones and the tones of Middle Chinese class I (33, 31) on the first syllable.

e. A rising tone on the second syllable combined with any low register tone except the low dipping tone on the first syllable will result in a rise-low level contour.

f. A rising tone on the second syllable combined with the high rising or the high falling tone on the first syllable will result in a high level-low level contour.

g. A low or high level tone on the second syllable combined with the high rising or high falling tone on the first syllable will result in a high level-low level contour.

h. A low rising tone on the first syllable combined with a high or low level tone on the second syllable will result in a continuously falling contour.

i. A low level tone on the first syllable combined with another low level tone or a high level tone on the second syllable will result in a rising-low level tone contour.

j. A high level tone on the second syllable combined with either of the two tones from Middle Chinese class I (33, 31) will result in a continuously falling contour.

k. A high falling tone on the second syllable combined with either of the two tones from Middle Chinese class I (33, 31) will result in a low level-mid level contour.

l. A low level tone on the second syllable combined with either of the two tones from Middle Chinese class I (33, 31) will result in a low level-rising contour.

m. A high level tone on the first syllable will result in a continuously falling contour when combined with a high rising tone on the second syllable, and in a high level-low level contour when combined with a low rising tone on the second syllable.

Table 2.5: Distribution of generalizations on the tone sandhi patterns in the speech of the young Wenzhou speakers.

<table>
<tr><th rowspan="2">First σ</th><th colspan="8">Second σ</th></tr>
<tr><th>33</th><th>31</th><th>35</th><th>24</th><th>42</th><th>11</th><th>313</th><th>212</th></tr>
<tr><td>33</td><td rowspan="2">18j</td><td rowspan="6">18c</td><td colspan="2">18m</td><td rowspan="2">18k</td><td rowspan="2">18l</td><td rowspan="8" colspan="2">18a</td></tr>
<tr><td>31</td><td colspan="2">18e</td></tr>
<tr><td>35</td><td>18g</td><td colspan="2">18f</td><td rowspan="4">18d</td><td>18g</td></tr>
<tr><td>24</td><td>18h</td><td colspan="2">18e</td><td>18h</td></tr>
<tr><td>42</td><td>18g</td><td colspan="2">18f</td><td>18g</td></tr>
<tr><td>11</td><td>18i</td><td colspan="2">18e</td><td>18i</td></tr>
<tr><td>313</td><td rowspan="2" colspan="5">18b</td></tr>
<tr><td>212</td></tr>
</table>

It seems that there is a certain strength ranking between the citation tones that decides which tone in the combination determines the pitch slope of the tone sandhi contour, and which tone can preserve some characteristics of the citation forms. It appears that the dipping tones are strongest in superimposing a uniform contour on the other tones when they are in final position, but weakest when in initial position, where they take on a low level slope while the second syllables retain their citation tones. The only exception to that are the low level and low falling tone, which are disambiguated from the dipping tone on the first syllable by the latter taking on a falling slope.

Further, the rising and falling tones superimpose left-prominent contours on any initial syllable they are combined with, with the exception of class I tones, which behave differently. The rising and falling tones on the second syllable transmit their characteristics (or end points) onto the first syllable, and take on a low level contour themselves. Only when the initial syllable is occupied by a class I tone do the second syllables take on a different tonal contour, either a falling or a high level pitch. The level tones are the odd ones out in most scenarios, and consequently, when they appear in second position, there is much more variability when it comes to the shape of the second tone.

Chapter 3

The tone sandhi domain of Wenzhou Chinese

3.1 Introduction

The research on tone sandhi, specifically on the type of tone sandhi that is found in many (southern) dialects of Chinese, has mainly been concerned with two types of research questions: (i) what is the correspondence between the lexical tones on the syllables and the tone sandhi contours, and how can it be explained phonologically and/or phonetically, and (ii) what is the domain of application for the tone sandhi changes, and which linguistic sub-domains affect its formation? This chapter aims to give an answer to the second question for the tone sandhi found in the speech of young speakers of Wenzhou Chinese.

3.1.1 Literature predictions on the tone sandhi domain in Wenzhou Chinese

The most clear-cut domain for the application of tone sandhi in Wenzhou Chinese is the disyllabic compound. Descriptions of the tone sandhi properties of Wenzhou either start with the tonal contours in disyllabic words (Chen 2000; Hou 1998; Qian 1992; Zhengzhang 1964a, 2008) or are even limited to this domain (Rose 2000, 2001, 2002, 2004). The compounds which are given as examples come from different word classes, encompassing nouns, verbs, adjectives, and adverbs (examples in (1)).[5]

(1)	Hanzi	Citation forms		Tone sandhi	Translation	Type
a.	天堂	*tʰi33*	*duɔ31*	42.31	‘paradise’	noun
b.	封口	*hoŋ33*	*kʰau35*	44.22	‘seal’	noun
c.	开始	*kʰi33*	*sz̩35*	44.22	‘to begin’	verb
d.	经理	*tɕaŋ33*	*lɛi24*	22.33	‘manager’	noun

[5] There is a general assumption in Chinese linguistics that an expression which is a compound in Standard Chinese will also be a compound in other dialects, going back to Chao 1968.

e.	干旱	*kɵ33*	*jø24*	22.33	‘dry’	adj.
f.	音乐	*jaŋ33*	*lu212*	35.31	‘music’	noun
g.	偏要	*p^hi33*	*ʔi212*	35.31	‘insistently’	adverb

Aside from disyllabic compounds, disyllabic structures of the type modifier-head or head-complement have also been described as tone sandhi application domains under certain conditions. Thus, both the lexicalized disyllabic adjective-noun constructions in (2a-b) and the more phrasal adjective-noun construction in (2c) are predicted to be treated as words in Wenzhou both prosodically and syntactically according to Chen 2000.

(2)	Hanzi	Citation forms		Tone sandhi	Translation
a.	热度	*ni212*	*dɵ11*	33.22	‘heat’ (lit. ‘hot degree’)
b.	白眼	*ba212*	*ŋa24*	22.35	‘disdain’(lit.‘white eye’)
c.	大树	*dou11*	*zʐ11*	35.22	‘big tree’

Other modifier-head constructions such as adverb-verb may also be treated like lexical words, or be realized with phrasal prosody, or even fluctuate between the two states. Chen (2000: 483) concludes that “lexicalization is a gradient process, and makes allowance for free variants”. He reaches a similar conclusion with respect to head-complement constructions, such as the verb-object constructions in (3), but remarks that “verb + object expressions [...] do not typically undergo [lexical tone sandhi]” (ibid). Thus, while there are some verb-object constructions which are lexicalized and therefore undergo tone sandhi, the majority is predicted to be realized with phrasal prosody.

(3)	Hanzi	Citation forms		Tone sandhi	Phrasal prosody	Translation
a.	打水	*tiɛ35*	*sʐ35*	44.22	11.35	‘to fetch water’
b.	开会	*k^hɛi33*	*vai212*	35.31	11.21	‘to atttend a meeting’
c.	请客	*t^hsɔŋ35*	*k^ha313*	35.42	11.31	‘to throw a party’

The “phrasal prosody” realization, as described by Chen, entails that the prosodically strong position retains its citation tone, whereas the prosodically weak position is “tonally reduced to zero “o”, phonetically interpreted as [a low tone].” (ibid). In compliance with the rule of “Tonic prominence” (Chen 2000:

500), the prosodically strong position is generally assumed by the syntactic non-head. This means that in the verb-object constructions in (3), the verb (as a syntactic head) is reduced to a low tone (11), whereas the object (as a syntactic non-head of the VP) assumes the prosodically strong position and is realized with its lexical tone in phrasal prosody.

The observation that disyllabic verb-object constructions sometimes may resist lexical tone sandhi has also been made for other dialects, such as Shanghai Chinese (Duanmu 1998), and relates to the general research question whether to classify verb-object constructions in Chinese as compounds or phrases (Chao 1968; Chi 1985; Dai 1998; Huang 1984). It is often agreed in Chinese linguistics that a single criterion for word- or phrasehood, such as syntactic mobility (*ba*-fronting) or the ability to take another object, is not sufficient to characterize all verb-object constructions. It is even the case that some disyllabic verb-object constructions might function as compounds in one context and as phrases in another (Huang 1984). Reversing the argument, it has even been proposed that the ability to undergo tone sandhi is the most stable criterion for wordhood of disyllabic verb-object constructions (Duanmu 1998).

What seems clear is that the status of verb-object constructions is difficult to define, and that there is a grey area in which specific examples may allow for both phrasal and lexicalized status. For this reason, the current chapter will experimentally investigate the ways in which young Wenzhou speakers realize verb-object constructions.

3.1.2 Influence of focus on the tone sandhi domain in Wenzhou Chinese

Outside of the syntactic status of specific words or phrases, it has also been proposed that "focus" may influence the formation of tone sandhi domains, and thereby "block" tone sandhi in contexts where it would apply outside the influence of focus. Specifically, in his description of Wenzhou, Chen 2000 describes that under the influence of focus, both the default prominence assignment and the default phrasing can be overridden. In his example (57), replicated as (4) below, the initial negation particle forms a prosodic domain of its own under focus in order to attract prosodic prominence (indicated by an asterisk).

(4)	not	eat	rice	
	BU	*chi*	*fan*	
a.	(	) (	*)	default reading
b.	(*)	(	) ( )	"focus" on BU

The claim that focus may "break up" tone sandhi domains and block tone sandhi has also been made for other dialects, such as Tianjin Chinese (Li & Liu 1985), Mandarin (Shen 1990a), and to a certain extent Shanghai Chinese (Selkirk & Shen 1990). For this reason, the current investigation will also test the influence of focus on the realization of disyllabic verb-object constructions in Wenzhou. If there is a focus effect on tone sandhi application, it can be expected that this effect should be even more clearly visible in disyllabic forms which naturally fluctuate between word and phrase status.

3.1.3 Current experiments

In order to separately test the two hypotheses that were put forward in previous research, two experiments were conducted. The first experiment was specifically concerned with the realization of disyllabic verb-object constructions with different degrees of lexicalization. The second experiment furthermore tested the influence of different contextual factors on the realization of disyllabic verb-object constructions, by putting them together in a list with disyllabic compounds. In addition, the recording condition of the stimuli in the second experiment was also varied between recording in isolation, in medial position in a carrier sentence, and in contrastive focus. In this way, it was attempted to exactly determine the influence of each of the contextual factors separately.

3.2 Experiment 1: Verb-object constructions and lexicalization

3.2.1 Stimuli

The first experiment was designed to specifically test whether the degree of lexicalization of a disyllabic verb-object construction correlates with its likelihood to be realized with tone sandhi contours in Wenzhou. For this purpose, 45 disyllabic verb-object constructions were first rated for their degree of lexicalization. The criteria used for the classification were taken from standard descriptions of Mandarin Chinese (Chao 1968; Li & Thompson 1981),

according to which a verb-object construction can be classified as a compound (i.e. lexicalized) if it fulfills any of the three following criteria.

(5) **Criteria for lexicalization of verb-object constructions**
a. One or both of the constituents [are] bound morphemes
b. Idiomaticity of the meaning of the entire unit
c. Inseparability or limited separability of the constituents

With the help of a native speaker and teacher of Chinese, the 45 verb-object constructions for experiment 1 were classified to fulfill none, one, or two of the lexicalization criteria in (5). None of the verb-object constructions were rated to fulfill all three criteria. The classification resulted in an almost even tripartite split of the examples, with 16 examples (35.6%) being classified as non-lexicalized, 13 examples (28.9%) being classified as lexicalized according to one of the criteria, and the remaining 16 examples (35.6%) being classified as lexicalized according to two of the criteria. The full list of examples and their lexicalization rating can be found in appendix 3.1.

3.2.2 Speakers

The speakers who participated in the two experiments were ten speakers (five male, five female) of the Wenzhou dialect of Chinese, aged between 18 and 20. They were high school graduates of the same high school in central Wenzhou and all born and raised in the central district of Lucheng in Wenzhou. None of them reported to have lived outside of Wenzhou for a significant amount of time within the last five years, and all of them considered themselves fluent speakers of the Wenzhou dialect. They were also fluent speakers of Standard Chinese, which they learned in school and were encouraged to use in conversations with each other on a daily basis. Of the ten speakers, three (two female, one male) participated in the elicitation experiment on verb-object constructions (experiment 1).

3.2.3 Experimental procedure

The recordings were made in a quiet recording studio in the TV and radio station in Wenzhou on an M-Audio Microtrack II portable digital recorder (44.1kHz, 16bits). The speakers were given a Sennheiser pc130 headset, and the position of the microphone was adjusted by the experimenter to ensure it was about 3 cm away from the corner of the mouth and outside of the immediate direction of aspiration. The speakers were seated at a table with about 50 cm distance to a

laptop screen (ACER TravelMate 280XCi), on which the stimulus sentences were presented in Chinese characters. All speakers confirmed they could read the sentences properly. The stimulus sentences were automatically randomized for every speaker and every trial, and presented one by one, with the experimenter determining the pace of succession.[6]

Before the start of the experiment, the speakers were told that they would see sentences on the computer screen, and asked to read them out aloud in Wenzhou dialect in a clear and natural way. If the speakers had long hesitation pauses within a sentence, they were asked to repeat the sentence, and if they spoke too softly, they were encouraged by the experimenter to speak up. All speakers were presented with the randomized stimulus sentences three times in a row, with a speaker-determined break in between trials. They were informed that they could interrupt or abort the experiment at any point. They received a small payment for their participation.

3.2.4 Data analysis

Before data analysis, any token which contained an error or mispronunciation was discarded. This concerned one token in experiment 1, and 22 tokens in experiment 2. The remaining tokens were analyzed acoustically and visually for their tone contour. Since the speakers themselves are not conscious of the tone sandhi process, the decision whether tone sandhi had taken place or not was based on the acoustic inspection of the tone contour. In both experiments, all stimuli were selected so as to ensure that this task would be reasonably straightforward. This means that the selected stimuli all had a tone sandhi contour that differed considerably from the citation tones on both syllables.

To decide whether tone sandhi had taken place or not, the recorded stimuli were compared to recordings of disyllabic compounds with the same combinations of citation tones. To illustrate the decision process, Figures 3.1 to 3.3 give examples of realizations of the stimuli in (6) by the same speakers as tone sandhi contour (left) and phrasal prosody (right).

(6)	Hanzi	Citation forms		Tone sandhi	Phrasal pr.	Translation
a.	跑马	*p^{h}uɔ35*	*mu24*	44.22	11.35	‘to ride a horse’
b.	坐车	*zuɔ24*	*ts^{h}u33*	35.22	11.33	‘to go by car’
c.	开门	*k^{h}ɛi33*	*maŋ31*	42.31	11.31	‘to open a door’

[6] The script that randomized and presented the stimuli was written by Jos Pacilly.

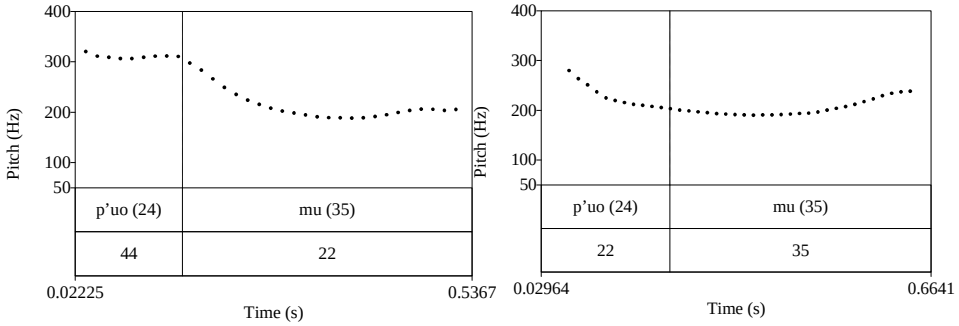


Figure 3.1: Realizations of stimulus (6a) by a female speaker.

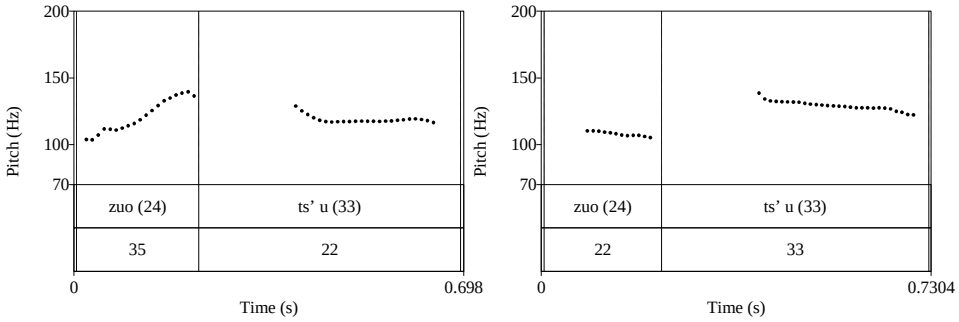


Figure 3.2: Realizations of stimulus (6b) by a male speaker.

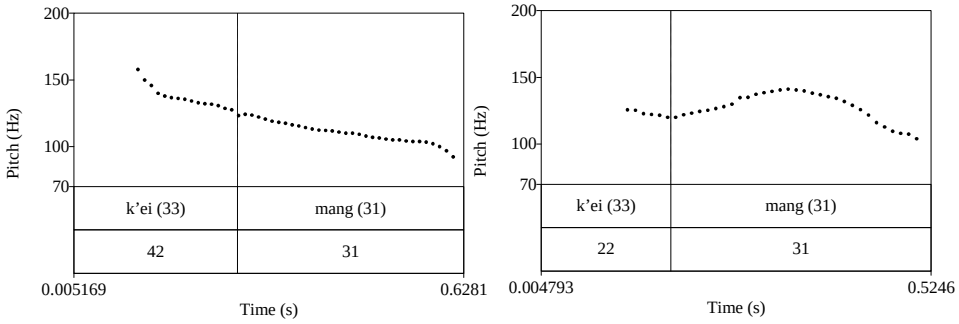


Figure 3.3: Realizations of stimulus (6c) by a male speaker.

3.2.5 Results

In the first experiment, a list of 45 verb-object constructions was read by three speakers three times. Acoustic inspection of the recorded tokens showed that the speakers varied the realization of the verb-object constructions not only between examples, but also sometimes produced the same example with tone sandhi contour in one recording, and with phrasal prosody in the next. For this reason, the realizations of the examples could not be averaged per speaker, but each realization entered as one token into the analysis.

Table 3.1 gives an overview of the observed realizations of the stimulus tokens, split by degree of lexicalization (0, 1, 2, according to the criteria laid out in section 2.1.1) and type of realization. Outside of the tone sandhi realization (TS) and the phrasal prosody realization (P), some tokens were also realized with a tonal contour that simply juxtaposed the citation tone values of the respective syllables. This type is classified as "other" in Table 3.1.[7]

Table 3.1: Absolute and relative (in parentheses) frequencies of realizations of verb-object constructions in experiment 1.

Degree of lexicalization	Observed contours			
	Phrasal prosody	Tone sandhi	Other	Total
0	123 (85.4%)	10 (6.9%)	11 (7.6%)	144
1	94 (80.3%)	17 (14.5%)	6 (5.1%)	117
2	90 (62.9%)	45 (31.5%)	8 (5.6%)	143
Total	307 (76.0%)	72 (17.8%)	25 (6.2%)	404

As can be seen in Table 3.1, there is indeed a connection between the degree of lexicalization and the tonal realization in the stimuli in experiment 1. More precisely, the speakers produced more tone sandhi contours on the lexicalized

[7] Since the overall token counts for the "other" contour are very low, it could be suspected that the inclusion of this type in the statistical analysis influenced the results of the comparison. For this reason, the statistical analysis was also performed on the dataset in which all "other" contours had been excluded. A Pearson Chi-square analysis still shows a statistically significant difference between the tone sandhi vs. phrasal prosody realizations in dependence on the degree of lexicalization [$\chi^2(2) = 30.4$, $p < 0.001$].

examples than on the non-lexicalized examples. Within the lexicalized examples, there was a further split between those examples that are lexicalized according to one of the criteria, and those that are lexicalized according to two of the criteria, with the latter being realized with tone sandhi contours more often than the former.

At the same time, it is true for all three types of examples that they are realized with phrasal prosody in the majority of cases. Even the examples that are clearly lexicalized receive phrasal prosody in 62.9% of the cases in experiment 1, and overall, more than three quarters of the tokens are realized with phrasal prosody, even though almost two thirds of the examples have been classified as "lexicalized". A Pearson Chi-square analysis confirms that this distribution is significantly different from chance [$\chi^2(4) = 31.18$, $p < 0.001$].[8]

3.3 Experiment 2: Verb-object constructions and context

3.3.1 Stimuli

A second experiment tested the influence of different contextual factors onto the realization of disyllabic verb-object constructions. Eight of the verb-object constructions from the first experiment were selected and alternated in a list with 34 disyllabic compounds, the latter of which are expected to be realized with tone sandhi contours in all cases. The eight verb-object examples in experiment 2 were classified into lexicalized and non-lexicalized according to the same (morphological, syntactic, and semantic) criteria used for the verb-object constructions in experiment 1. The division between lexicalized and non-lexicalized verb-object constructions turned out to be similar to that of the stimuli in experiment 1, namely 37.5% non-lexicalized examples (three of eight), and 62.5% lexicalized examples (five of eight). Of the lexicalized examples, three (37.5%) were lexicalized according to one of the criteria outlined in (5), and two (25%) were lexicalized according to two of the criteria.

The complete stimulus list was presented to the speakers in three different conditions: the disyllabic forms were elicited (i) in isolation, (ii) in

[8] Technically, Chi-square analysis is not possible for the data presented here, since this test is based on the assumption of independence between observations, which is not warranted in the current dataset (since multiple tokens were collected from the same individual). Nonetheless, this test allows for categorical variables, and therefore it can still give a useful indication for the realiability of the counts.

medial position in a carrier sentence as in (7), and (iii) in medial and final position in a carrier sentence which induced contrastive focus, as in (8).

(7) 我 是 说 TARGET 1 这个 词.
ŋ zɻ kuɔ TARGET 1 *kai zɻ*
I ASP say TARGET 1 this word
'I said the word TARGET 1.'

(8) 我不是 说 TARGET 1 这个词, 我是 说 TARGET 2.
ŋ nau zɻ kuɔ TARGET 1 *kai zɻ ŋ zɻ kuɔ* TARGET 2
I not ASP say TARGET 1 this word I ASP say TARGET 2
'I did not say the word TARGET 1, I said TARGET 2.'

In the last condition, the verb-object constructions and disyllabic compounds were paired such that two segmentally and tonally similar stimuli would appear in the same sentence, in order to further encourage the speakers to differentiate the stimuli by use of contrastive prosody. The order of the paired items was varied per sentence, so that each speaker saw the same item twice per list, once in medial and once in final position. A full list of the stimulus pairs of experiment 2 can be found in appendix 3.2.

3.3.2 Speakers, experimental procedure, and data analysis

Of the ten speakers introduced in section 3.2.2, two speakers recorded the stimuli of experiment 2 in isolation (one male, one female) and two in medial position in a carrier sentence (one male, one female). All ten speakers, including the three speakers from experiment 1, participated in the elicitation of the stimuli of experiment 2 in contrastive focus carrier sentences. The experimental procedure and data analysis were identical to those of the first experiment, as outlined in sections 3.2.3 and 3.2.4.

3.3.3 Results

In the second experiment, a list containing eight verb-object constructions and 34 disyllabic compounds was read by different groups of speakers in isolation, in medial position in a carrier sentence, and in a carrier sentence inducing contrastive focus. First of all, an analysis of the prosodic realizations of the disyllabic compounds showed that they were realized with tone sandhi contours in the vast majority of cases. In the cases when they were not realized with tone

sandhi contours, the speakers either produced the citation tones on both syllables, or adopted a realization strategy which was clearly different from the tone sandhi prosody observed in the verb-object constructions in both experiments. These cases are counted as "other" in Table 3.2.

Table 3.2 gives an overview over the realizations of the disyllabic compounds in experiment 2, split by three positional conditions: individual (isolation + medial position in a carrier sentence), medial (medial position in a contrastive focus carrier sentence), and final (final position in a contrastive focus carrier sentence). Statistical analysis shows that there is no statistically significant difference between the realization ratios in the three different conditions [$\chi^2(2) = 4.6$, $p = 0.1$, ns].

Table 3.2: Absolute and relative (in parentheses) frequencies of realizations of disyllabic compounds in experiment 2, compared between different recording situations.

Position	Observed contours		
	Tone sandhi	Other	Total
Individual	381 (96.2%)	15 (3.8%)	396
Medial	970 (96.4%)	36 (3.6%)	1006
Final	955 (94.6%)	55 (5.4%)	1010
Total	2306 (95.6%)	106 (4.4%)	2412

The eight verb-object constructions of experiment 2 were analyzed separately for their realizations in the three recording situations, and the realizations were compared to the degree of lexicalization of the examples. Tables 3.3, 3.4, and 3.5 show the results for the verb-object constructions of experiment 2 in isolation (Table 3.3), in medial position in a carrier sentence (Table 3.4), and in a contrastive focus sentence (Table 3.5).

Table 3.3: Absolute and relative (in parentheses) frequencies of realizations of verb-object constructions in experiment 2 in isolation.

Degree of lexicalization	Observed contours		
	Phrasal prosody	Tone sandhi	Total
0	4 (22.2%)	14 (77.8%)	18
1	0 (.0%)	17 (100%)	17
2	0 (.0%)	12 (100%)	12
Total	4 (8.5%)	43 (91.5%)	47

Table 3.4: Absolute and relative (in parentheses) frequencies of realizations of verb-object constructions in experiment 2 in a carrier sentence.

Degree of lexicalization	Observed contours		
	Phrasal prosody	Tone sandhi	Total
0	5 (27.8%)	13 (72.2%)	18
1	0 (.0%)	12 (100%)	12
2	0 (.0%)	12 (100%)	12
Total	5 (11.9%)	37 (88.1%)	42

Table 3.5: Absolute and relative (in parentheses) frequencies of realizations of verb-object constructions in experiment 2 in a carrier sentence inducing contrastive focus.

Degree of lexicalization	Observed contours			
	Phrasal prosody	Tone sandhi	Other	Total
0	37 (20.8%)	133 (74.7%)	8 (4.5%)	178
1	0 (.0 %)	165 (98.8%)	2 (1.2%)	167
2	19 (15.8%)	101 (84.2%)	0 (.0%)	120
Total	56 (12.0%)	399 (85.8%)	10 (2.2%)	465

As can be seen in the above tables, there is a clear trend in all three recording contexts to have more phrasal prosody realizations for the non-lexicalized examples (0) than for the lexicalized examples (1, 2), and more tone sandhi in the lexicalized compared to the non-lexicalized examples. The distributions are significantly different from chance for all recording situations: isolation [$\chi^2(2) = 7.04$, $p = 0.03$], carrier sentence [$\chi^2(2) = 7.57$, $p = 0.023$], and contrastive focus [$\chi^2(4) = 46.55$, $p < 0.001$].[9]

At the same time, when pooling over the different degrees of lexicalization and comparing only the overall amount of realizations between recording situations, the distributions are similar in all three recording situations. Statistical analysis confirms that there is no significant difference between recording situations [$\chi^2(4) = 2.54$, $p = 0.637$, ns], as can be seen in Table 3.6.[10]

Table 3.6: Absolute and relative (in parentheses) frequencies of realizations of verb-object constructions in experiment 2, compared between different recording situations.

Recording situation	Observed contours			
	Phrasal prosody	Tone sandhi	Other	Total
Contrastive focus	56 (12.0%)	399 (85.8%)	10 (2.2%)	465
Carrier sentence	5 (11.9%)	37 (88.1%)	0 (.0%)	42
Isolation	4 (8.5%)	43 (91.5%)	0 (.0%)	47
Total	65 (11.7%)	479 (86.5%)	10 (1.8%)	554

Furthermore, there is no significant difference in the observed realizations when comparing the occurrences of the verb-object constructions in medial position in the contrastive focus carrier sentences with those in final position in contrastive

[9] For the statistical analysis of the contrastive focus carrier sentence, again the presence of the category "other" does not change the statistical result much. In a comparison of only the tone sandhi vs. phrasal prosody realizations, the Pearson Chi-square analysis still returns a statistically significant difference for the different degrees of lexicalization [$\chi^2(2) = 38.63$, $p < 0.001$].

[10] Again, excluding the "other" condition does not change the statistical results much [$\chi^2(2) = 0.58$, $p = 0.747$, ns].

focus carrier sentences. Pooled over the different degrees of lexicalization, the amount of tokens that are realized with tone sandhi or phrasal prosody is not statistically significantly different from chance [$\chi^2(2) = 1.99$, $p = 0.37$, ns], as can be seen in Table 3.7.[11]

Table 3.7: Absolute and relative (in parentheses) frequencies of realizations of verb-object constructions in experiment 2 in contrastive focus carrier sentence , compared for medial vs. final position.

Position in focus carrier sentence	Observed contours			
	Phrasal prosody	Tone sandhi	Other	Total
Final	33 (14.2%)	195 (83.7%)	5 (2.1%)	233
Medial	23 (9.9%)	204 (87.9%)	5 (2.2%)	232
Total	56 (12%)	399 (85.8%)	10 (2.2%)	465

3.4 Discussion

To account for the realization of verb-object constructions in tone sandhi languages like Wenzhou Chinese, two predicting factors have been put forward in the literature. It was proposed that a disyllabic verb-object construction is more likely to undergo tone sandhi and to be treated like a compound if it is lexicalized, and more likely to be realized with phrasal prosody if it is not lexicalized or subject to focus.

The current experiment confirms the first hypothesis: in both experiments described above, a significant connection was found between the number of tokens that were realized with a tone sandhi contour by the speakers, and the classification of these tokens as lexicalized according to grammatical factors (syntax, semantics, morphology). For the second factor of focus on the other hand, the prediction was not borne out. The ratio of tokens realized with phrasal prosody vs. tone sandhi prosody was similar in isolation and in a condition that induced contrastive focus on the target stimuli.

[11] Again, excluding the "other" condition does not change the statistical results much [$\chi^2(1) = 1.99$, $p = 0.159$, ns].

However, the results presented in the previous section show that the relative number of tone sandhi vs. phrasal prosody realizations differed substantially between the two experiments, even though the list of verb-object construction stimuli in the two experiments was similar with respect to the factor of lexicalization. Nonetheless, it can be observed that when the speakers are presented with a list of stimuli which exclusively contains verb-object constructions, they predominantly realize them with phrasal prosody, and even the clearly lexicalized examples receive phrasal prosody in the vast majority of cases (71.6% on average).

On the other hand, if the speakers are presented with a list which contains some verb-object constructions, but predominantly consists of disyllabic compounds, the speakers use the same tone sandhi realization for the verb-object constructions that they use for disyllabic compounds in the majority of cases, irrespective of the degree of lexicalization of the verb-object construction examples (86.5% on average). This is true regardless of the recording context (isolation versus carrier sentence), the presence versus absence of contrastive focus, and the position of the target word within the contrastive focus carrier sentence (medial versus final).

While lexicalization therefore has some influence on the realizations of verb-object constructions by the young Wenzhou speakers, accounting for an average of 20% difference between the lexicalized and the non-lexicalized examples, it seems that the composition of the stimulus list, and thereby the prosodic context in which the verb-object constructions appear, has a much greater influence on how these verb-object constructions are realized by the speakers (71.6% phrasal prosody in verb-object context vs. 86.5% tone sandhi in lexical compound context).

In order to show that this effect is not dependent on the specific speakers recorded in the two experiments, or on the size of the speaker pool which was recorded for each experiment, Table 3.8 compares the realizations of the verb-object constructions in experiment 2 between two speaker groups. The first group (1) consists of the three speakers which were also recorded for experiment 1, and the second group (2) consists of the remaining seven speakers who were only recorded for experiment 2. If the speakers who recorded experiment 1 are overall more likely to realize any verb-object construction with phrasal prosody rather than with tone sandhi contours, their realizations of the verb-object constructions in experiment 2 should differ markedly from the realizations by the other seven participants.

A statistical analysis on the findings in Table 3.8 shows that this prediction is not borne out. There is no statistically significant difference between the relative number of tone sandhi vs. phrasal realizations of the stimuli of experiment 2 by the speakers of experiment 1, compared to the remaining seven speakers of experiment 2 [$\chi^2(2) = 4.79$, $p = 0.09$, ns].[12]

Table 3.8: Absolute and relative (in parentheses) frequencies of realizations of verb-object constructions in experiment 2 pooled over all recording conditions, compared between the three participants of experiment 1 and the remaining seven participants.

Speaker group	Observed contours			
	Phrasal prosody	Tone sandhi	Other	Total
Experiment 1 & experiment 2	27 (14.1%)	158 (82.7%)	6 (3.1%)	191
Experiment 2 only	38 (10.5%)	321(88.4%)	4 (1.1%)	363
Total	65 (11.7%)	479 (86.5%)	10 (1.8%)	554

The findings of this study suggest that the variability in the prosodic realization of verb-object constructions is even greater than previously assumed, and that it not only depends on the exact grammatical function that a verb-object construction plays in a given sentence, but also on the prosodic context in which it is uttered. It appears that in recording situations in which speakers are applying tone sandhi to a list of compounds, they are much more likely to extend this tone sandhi application to disyllabic constructions which are not lexically compounded. In contrast, in a context where all stimuli share a certain grammatical analysis of their components, speakers are more likely to be biased by this analysis in their prosodic realization of the examples, and less likely to just treat them like any other disyllabic structure.

[12] Again, excluding the "other" condition does not change the statistical results much [$\chi^2(1) = 1.87$, $p = 0.172$, ns].

3.5 Conclusion

This study set out to investigate two claims which have been put forward in the discussion concerning the tone sandhi domain of Wenzhou Chinese. The first claim, which could be confirmed in this study, predicted that disyllabic constructions which are grammatically ambiguous between word and phrase status will also be treated ambiguously in their prosodic behavior. More specifically, it was found that the degree of lexicalization of verb-object constructions in Wenzhou Chinese according to semantic, syntactic, and morphological criteria was significantly correlated with the likelihood of the verb-object construction receiving a tone sandhi realization. If a verb-object construction was lexicalized, it was realized with tone sandhi contours more often, and thereby treated alike to lexical compounds, than if it was not lexicalized.

The second prediction from the literature, however, could not be confirmed by the dataset presented in this study. In the data recorded here, the presence of contrastive focus on the target stimuli had no statistically significant effect on the speakers' choice of realization of the verb-object constructions as compounds or as phrases. In this light, the prediction that focus would induce a prosodic boundary which would in turn block the application of tone sandhi was not borne out. Rather, the same set of stimuli was (on average) realized with a comparable ratio of tone sandhi and phrasal prosody realizations in isolation and under focus.

However, a comparison of the results of the two experiments reported in this study showed that another factor significantly affected the realizations of the verb-object constructions. As it turned out, the verb-object constructions in experiment 2, which were recorded in a list that majorly contained disyllabic compounds, were realized with tone sandhi prosody in most cases, whereas in a comparable list of verb-object constructions with no disyllabic compounds interspersed, the examples were realized predominantly with phrasal prosody. Since the ratio of lexicalized and non-lexicalized verb-object constructions was similar in both lists, it appears that it is most likely the prosodic context itself which biased speakers towards treating the verb-object constructions as disyllabic lexemes in the context of disyllabic compounds.

These findings show that for the young speakers of Wenzhou recorded here, it is indeed true that their realizations of verb-object constructions are highly variable and contextually dependent. At the same time, it could be shown that the contextual factors which play the most important role for their

realization, are different from what has been assumed in the literature so far. While lexicalization plays a partial role for the prosodic realization of the verb-object construction, the observed variability appears to be dependent on the prosodic context in which a specific construction is uttered.

Chapter 4

The effect of the tone sandhi domain on focus expression in Wenzhou Chinese

4.1 Introduction

This chapter is concerned with the phonetic marking of contrastive focus, and the influence of prosodic structure on this phonetic marking. For intonation languages, it has been found that unlike other types of focus, contrastive focus can be marked not only on words, but even on individual syllables of words, and (for proficient speakers) even on parts of syllables (Sluijter 1992; Sluijter & van Heuven 1995; van Heuven 1994). In such cases, speakers lend additional prominence to the contrasted parts of the word, using phonetic markers such as F_0/pitch accentuation, lengthening, intensity, and spectral distribution. The examples in (1) illustrate this for English, with the contrasted parts marked in bold and upper case.

(1) a. I said '**COF**fin', not '**MUF**fin'.
b. I said '**P**it', not '**B**it'.
c. I said 'cof**FIN**', not 'cof**FEE**'.

For Dutch, it has been found that in examples comparable to (1c), where the focus contrast is located on a syllable which is lexically unstressed, the pitch accent which normally marks stress is "shifted" to the unstressed syllable by virtue of focus marking. At the same time, the accompanying durational marking of stress is only partially affected by focus: the focused (lexically unstressed) syllable is lengthened, but this lengthening does not fully reverse the durational difference between the two syllables of the word (Sluijter & van Heuven 1995).

These findings suggest that in intonation languages like English, where pitch accentuation serves a dual function both as a primary marker of lexical stress and as a primary marker of focus, focus can trump phonology and determine the location of the pitch accent within a word. At the same time, it appears that duration as a secondary marker for both focus and lexical stress is much less volatile than pitch accentuation: a lexically unstressed syllable is lengthened under focus, but its relative duration in the word is still shorter than

that of a lexically stressed syllable. Apparently, focus cannot fully override the durational marking of stress, so that a residue of durational stress marking remains intact even under contrastive focus on a non-stressed syllable.

The current study looks at Wenzhou Chinese, a tone language with an extensive tonal phonology. In this language (as in many related dialects of the Wu dialect group), disyllabic words represent an important domain for phonological processes, because when two syllables come together as a compound, the lexical tone on both syllables changes in a process called "tone sandhi". In that sense, it can be said that speakers use the tone change to mark for listeners that the two syllables form one compounded constituent, rather than just two adjacent syllables. Therefore, the tone change process serves as a marker for wordhood. This chapter reports the results of an experiment which tests whether contrastive focus can be marked within this tone sandhi domain, at the expense of the phonological marking of wordhood.

4.1.1 Focus marking in Chinese

Across different Chinese dialects and focus domains of differing sizes, two phonetic correlates of focus have been in the center of attention: F_0 and duration modification.[13] However, most of the studies were concerned with the effect of focus on lexical tones in sentences, whereas the current study investigates tones that result from tone sandhi (cf. section 4.1.3). Furthermore, most studies only considered entire lexical words as focus domains, without testing what happens to the focus marking if only a part of a word is focused. The current study aims to combine these two fields of exploration, and investigate focus marking below the word level on sandhi tones.

Of the two phonetic effects that have been identified as correlates of focus in Chinese, F_0 is the more straightforward one: all Chinese dialects are tonal, and differences in the implementation of these tones between focus and non-focus conditions are readily observable in most dialects. Following the detailed exploration of focus effects in Standard Chinese by Xu (1999), F_0 differences between focus and non-focus condition have been reported for several dialects of Chinese, such as Mandarin Chinese (Wang & Xu 2006),

[13] The experiment reported in this chapter specifically investigates the effect of contrastive focus. However, in the discussion of previous findings on focus effects in Chinese, other types of focus, such as focus induced by answering a wh-question, have been included in the literature review. The assumption that both types of focus elicit comparable effects is interesting in itself, but its verification lies outside of the scope of this study.

Cantonese (Gu & Lee 2007b), Shanghai Chinese (Chen 2009), and Taiwan Mandarin (Chen et al. 2009; Xu et al. 2012). Generally, the reported findings can be summarized as F_0 range expansion under focus, such that the F_0 maximum will be higher, and the F_0 minimum will be lower under focus than in the non-focus control conditions.

In most, but not all of these dialects, the on-focus F_0 expansion effect of focus is accompanied by a F_0 compression effect on the post-focal stretch of the target sentence, which can represent an enhancing cue for listeners to determine the location of the focused constituent in the sentence (Chen et al. 2009; Xu et al. 2012). However, taking a closer look at the phonetic nature of "post-focus compression" in Standard Chinese, Chen (2010) reports that the tone contours on post-focal tones differ depending on the preceding tone. Rather than a uniform lowering effect in post-focal condition, different tonal contexts share a weak tonal implementation in post-focal position, which manifests itself in hypoarticulation of the tonal target, a reduced degree of distinctiveness of the tonal contour, and a greater influence of the preceding tones, especially if the latter were in on-focus condition themselves.

While F_0 manipulation may be a widespread phonetic means, it is not a necessary focus marker across all Chinese dialects. As evident in studies on Hong Kong Cantonese (Wu & Xu 2010) and Taiwan Min (Chen et al. 2009; Xu et al. 2012), in some dialects, the implementation of tonal contours is not systematically influenced by the presence or absence of focus. Rather, in these dialects, duration and intensity of focused syllables seem to play a greater role in signaling focus.

Outside of F_0 manipulation, duration manipulation has been reported as the other stable cue to different focus conditions in Chinese dialects. Lengthening of new or contrastively focused syllables has been found in Taiwan Mandarin (Pan et al. 2005), Taiwan Min (Chen et al. 2009; Xu et al. 2012), Cantonese (Gu & Lee 2007b), Shanghai dialect (Chen 2009), and Hong Kong Cantonese (Wu & Xu 2010). Interestingly, even dialects of Chinese that do not employ F_0 manipulation for focus marking show lengthening of the syllables that are under focus.

Studies like Xu 1999 and Chen & Gussenhoven 2008 have shown for Standard Chinese that there is a stable lengthening effect within the focused word, whereas the durations hardly show any difference between neutral (no-focus), pre-focus, and post-focus condition. In other words, lengthening seems to be more local to the focused constituent, compared to the more global effect of F_0 manipulation. Furthermore, while the shape and height of F_0 contours is

influenced by a number of factors outside of focus (such as sentence position/downtrend within the sentence and the shape/height of surrounding tones), the durational influence of focus appears to be stable across different sentence positions of the focused word.

4.1.2 Focus effects and prosodic domains in Chinese

Most of the results described in the previous two sections are derived from studies that investigate focus on the word level, usually with the domain of focus being either monosyllabic or disyllabic words. However, a couple of recent studies have also asked how the focus effect is applied in longer words, or in words that serve as application domain for phonological processes, such as tone sandhi.

In Chen 2006, focus effects on the durational distribution in quadrisyllabic words in Standard Chinese were tested, including focus domains smaller than the whole word (either the disyllabic foot or the syllable). If the focus domain was smaller than the word, lengthening targeted the focus domain, and to a certain extent the immediately adjacent syllables. For example, focus on just the initial syllable of the quadrisyllabic word resulted in lengthening of that syllable, but the second syllable was also lengthened through "rightward spill-over lengthening" (Chen 2006: 197). This shows that when a focus domain undercuts a morphophonological domain, the lengthening effect may exceed the immediate focus domain.

Furthermore, the results have shown that focal lengthening applies unevenly if there is an inherent duration difference present within the focus domain. For example, in Shanghai Chinese, syllables can be intrinsically short if they end in a glottal stop or a nasal coda (Chen 2008). Under focus on disyllabic words with either inherently short or inherently long first syllables, Chen (2009) found that in words with an intrinsically long initial syllable, both syllables are lengthened under focus. In contrast, in words with an intrinsically short initial syllable, the first syllable was lengthened very little, and the second syllable received a more pronounced lengthening. This suggests some form of "compensatory lengthening" (Chen 2009), and can even be interpreted as evidence for the assumption "that durational modification for focus may be computed over the whole bi-syllabic sandhi domain" (ibid).

For F_0 effects in a focus domain below the word level, the only reported results come from Cantonese (Gu & Lee 2007b). In this study, the speaker was given disyllabic nonsense words in three different focus conditions: no focus, narrow focus on the first syllable, and narrow focus on the second syllable. The

authors report that the on-focus F_0 expansion effect of focus starts slightly before the narrowly focused syllable and decreases gradually over time, but by and large, it seems that the Cantonese speaker is able to focus only one of the syllables of the word. However, since the tested words were nonsense words, and since disyllabic compounds do not receive any special prosodic marking in Cantonese, it can still be asked whether a similar result can be found for a different dialect when testing actual compound words.

4.1.3 Wenzhou Chinese

Compared to many other dialects of Chinese, the disyllabic word domain in Wenzhou has a special status, since it serves as the application domain for tone sandhi. This means that when two syllables come together in a disyllabic compound, the tones on both syllables change in a regular, but not immediately transparent way. In that sense, the tone sandhi contour on the two syllables is functionally loaded, since it signals to the listener both information about the original lexical tones on the two syllables, and the fact that the two syllables have been compounded into a disyllabic word. If the focus effect turns out to be different in Wenzhou, this would indicate a sensitivity of focus marking to prosodic domainhood.

In some traditional impressionistic accounts of focus and tone sandhi, it has been reported that focus can block tone sandhi, for example by inserting a prosodic boundary before or after the focused constituent (Selkirk & Shen 1990; Shih 1997). Such a tone sandhi-blocking effect of focus has also been hinted at for Wenzhou Chinese (Chen 2000). However, for the Wenzhou dialect as spoken by the younger speakers today, it can be observed that tone sandhi in disyllabic words also applies under focus, even if the focus domain undercuts the tone sandhi domain. The interest of the present study is therefore to investigate whether the tonal contour that results from the application of tone sandhi still reflects the location of contrastive focus phonetically.

4.1.4 Current experiment

The current experiment is intended to investigate the effect of mismatch between prosodic domains and focus domains. The prosodic domain of interest in the current experiment is the disyllabic word, which serves as the domain for phonological tone sandhi processes in the dialect of Wenzhou Chinese. It will be systematically combined with narrow focus on five different locations with respect to the disyllabic target word.

- S1 focus: Focus on the first syllable of the disyllabic target word
- S2 focus: Focus on the second syllable of the disyllabic target word
- Word focus: Focus on the whole disyllabic target word
- Pre-target focus: Focus on the word preceding the disyllabic target word
- Post-target focus: Focus on the word following the disyllabic target word

By comparing the realizations of the tone sandhi contour in the different focus conditions, it will be investigated whether Wenzhou speakers prioritize the marking of the precise focus location, as they do by shifting the pitch accent in intonation languages, or whether the preservation of the sandhi contour as a marker of compounding will take precedence.

4.2 Methods

4.2.1 Stimuli

In order to limit the investigation of the effect of focus on a specific tone combination in Wenzhou, only disyllabic target words with a rising-falling tone sandhi contour were tested. The rise-fall tone sandhi contour results from the combination of any tone with one of the two dipping tones of the language. In this combination, any citation tone on the initial syllable will become rising, and the dipping tone on the second syllable will become falling. In contrast to level tones, contour tones (such as falling and rising tones) can be expanded in their F_0 span in both directions (i.e. upwards and downwards), and are therefore well suited for the investigation of F_0 effects of focus.

Apart from tone sandhi, the Wenzhou dialect also displays the division of the lexical tones into "registers". Broadly speaking, each tonal contour of Wenzhou (level, rising, dipping, falling) can manifest itself either in a "low" (L) or in a "high" (H) register form, co-varying with the voicing properties of the onset of the tone-bearing syllable. In order to control for register effects, all four possible combinations of registers were included in the current experiment: high register on the first and on the second syllable (HH), high register on the first and low register on the second syllable (HL), low register on the first and high register on the second syllable (LH), and low register on both syllables (LL). One of the examples that was classified as (LL beforehand was realized by the speakers as (LH), which means that there were four examples in the category (LH), two in the category (LL), and three each in the categories (HH) and (HL). All examples were compounds of the structure (noun-noun).

The tested stimulus words are listed in (2) with the following information from left to right: the Chinese characters as the speakers saw them during the experiment, a broad transcription of the target words in Wenzhou dialect, the citation tones on the first and second syllable, and a translation into English. The Chao numbers describing the citation tones are based on the description of Wenzhou in You 2002.

(2)	Register	Hanzi	Citation forms		Translation
a.	HH	中国	*tɕoŋ33*	*kai313*	'China'
b.	HH	战国	*tɕi42*	*kai313*	'Warring states period'
c.	HH	教室	*kuɔ42*	*sai313*	'classroom'
d.	HL	短袜	*tɵ35*	*mu212*	'socks'
e.	HL	中学	*tɕoŋ33*	*ɦu212*	'middle school'
f.	HL	小麦	*ɕiɛ35*	*ma212*	'wheat'
g.	LH	牙刷	*ŋu31*	*sɵ313*	'toothbrush'
h.	LH	外国	*va11*	*kai313*	'foreign country'
i.	LH	蜡烛	*ɦɛi212*	*sɛi313*	'candle'
j.	LH	语法	*nø24*	*hu313*	'grammar'
k.	LL	文学	*vaŋ31*	*ɦu212*	'literature'
l.	LL	腊肉	*la212*	*ɳɒu212*	'bacon'

The target words were embedded in a carrier sentence that remained constant across the five focus conditions, as illustrated with one of the examples in (3). This carrier sentence was coupled with a question which induced contrastive focus in one of five locations within the carrier sentence, by prompting the speaker to correct a part of the question in the answer sentence. All question-answer combinations were checked by a native speaker of Wenzhou and three native speakers of other Chinese dialects prior to the experiment to ensure grammaticality and naturalness of the sentences. In the following example, the contrasted part is marked in bold in the context questions for the reader's benefit, but it was not typographically marked for the speaker during the recording in any way.

(3) **Target sentence:**

a.	不，	我说	牙刷	商店	这几个字眼.
	fu	*ŋ kuɔ*	*ŋu sɵ*	*ɕi ti*	*ki ki kai zɻ ŋa*
	no	I say	TARGET	FRAME	these words

'No, I say the words toothbrush shop.'

S1 focus:

b.	你	说	**鞋**刷	商店	啊？
	ni	*kuɔ*	***ɦa*** *sɵ*	*ɕi ti*	*a?*
	you	say	TARGET	FRAME	Q

'Are you saying **shoe** brush shop?'

S2 focus:

c.	你	说	牙**膏**	商店	啊？
	ni	*kuɔ*	***ŋu kɜ***	*ɕi ti*	*a?*
	you	say	TARGET	FRAME	Q

'Are you saying tooth**paste** shop?'

Word focus:

d.	你	说	**短袜**	商店	啊？
	ni	*kuɔ*	***tɵ mu***	*ɕi ti*	*a?*
	you	say	TARGET	FRAME	Q

'Are you saying **sock** shop?'

Pre-target focus:

e.	你	**写**	牙刷	商店	啊？
	ni	***ɕi***	*ŋu sɵ*	*ɕi ti*	*a?*
	you	write	TARGET	FRAME	Q

'Are you **writing** toothbrush shop?'

Post-target focus:

f.	你	说	牙刷	**容器**	啊？
	ni	*kuɔ*	*ŋu sɵ*	***joŋ*** *ts^hɻ*	*a?*
	you	say	TARGET	FRAME	Q

'Are you saying toothbrush **container**?'

The disyllabic target phrase itself was part of a quadrisyllabic phrasal construction, and syntactically acted as modifier for the following disyllabic compound (glossed as FRAME in (3)). In an elicitation prior to the pilot experiment, it was ensured that in a quadrisyllabic phrasal construction like the one used in the experiment, the initial disyllabic construction would have a

clearly visible rising-falling tone contour, and not be tonally reduced due to its modifier status. The complete list of examples and precursor questions can be found in the appendix at the end of the thesis (appendix 4.1).

4.2.2 Speakers

The speakers recorded for the current experiment were all between 20 and 29 years of age (mean age = 24:0). They were all born and raised in the inner-city Lucheng district of Wenzhou, and spoke the local dialect with their friends and family on a regular basis. All of them were fluent in Standard Chinese, but had no difficulty reading out aloud Chinese characters in their dialect. None reported any hearing or speech impediments. All in all, 18 speakers were recorded for the current experiment. The data from three speakers had to be excluded from the analysis because of excessive numbers of errors and hesitations in their data, leaving the data from 15 speakers for analysis (twelve female). Five of them recorded the materials once, while the remaining ten speakers were recorded twice.

4.2.3 Experimental procedure

Speakers were recorded in a sound-proofed recording studio in Wenzhou in individual sessions, and received a small payment for their participation. Each speaker was seated in front of a 13" monitor and given a Sennheiser pc130 headset. The experimenter ensured that the microphone of the headset was placed approximately 3 cm from the corner of the mouth of the subject. Via an external digitizer (UA-G1), the sound was recorded directly on the laptop (Acer Aspire 1810TZ) on which the stimuli were displayed to the subject.

The speakers were first informed about the recording procedure. They were instructed to read out phrases and sentences presented on the screen using Wenzhou dialect in a natural and clear manner. If they were unsure how to pronounce a word or phrase, they could skip to the next item, and if they felt they had made a mistake, they could go back and repeat the recording of the previous item. They were told that they could interrupt or abort the recording at any point.

The recording itself was done using a script in the computer program PRAAT (Boersma & Weenink 2001).[14] This script presented the stimuli one by one, and recorded each stimulus individually after the speaker initiated the

[14] The script used for presenting and recording the stimuli was written by Jos Pacilly, and slightly modified by the author.

recording. Before the actual recording, all speakers completed a practice series with eight short phrases that were not part of the actual experiment. This was done in order to familiarize the speakers with the self-managed recording procedure. After completing the practice items, the speakers were asked to indicate whether they understood the recording procedure and were ready to start the actual experiment.

4.2.4 Data analysis

After the recording, all utterances were checked for mistakes and hesitations. Any pause longer than 100 ms within the carrier sentence was counted as a hesitation, and the relevant recordings were excluded from further analysis. This resulted in a total of 1424 recorded sentences which entered the analysis. All sentences were manually segmented by marking the beginning and end of the sentence, target words, and the rhymes of the target syllables, using acoustic and visual inspection of the sound wave and spectrogram. A PRAAT script measured and extracted the duration of each syllable of the target word, and performed F_0 measurements at 20 equidistant intervals within the rhymes of the target word syllables.[15] Before F_0 extraction, the measurements were checked for octave jumps and tracking anomalies due to creaky voice, and manually corrected (one octave up or down) where necessary (69 cases out of 1424 tokens).

4.3 Results

In order to assess the influence of the different focus conditions on the realization of the tone sandhi contour in the target word, several Repeated Measures (RM) ANOVAs were performed with subjects ($F1$) and items ($F2$) as random factors. Post-hoc pairwise comparisons were performed to investigate significant differences between the focus conditions, using the Sidak adjustment for multiple comparisons. The significance level adopted for the post-hoc comparisons was $p < 0.01$. All reported degrees of freedom have been Huyhn-Feldt corrected when the requirement of sphericity was not met.

[15] The script used for segmenting and measuring the files was written by Jos Pacilly.

4.3.1 F_0 effects

4.3.1.1 Register effects on F_0

In order to statistically determine whether the register on either syllable (S1, S2) interacted with the focus effects, the F_0 range on both of the two target syllables was calculated using the formula 12*ln(Fmax/Fmin)/ln(2). A by-subjects RM ANOVA on the so-calculated range as dependent variable was performed, testing the two dependent variables FOCUS (five levels) and REGISTER (four levels). This resulted in a significant main effect for FOCUS [S1 range: $F(1.28,14) = 24.74$, $p < 0.001$, S2 range: $F(2.16,14) = 38.56$, $p < 0.001$] and for REGISTER [S1 range: $F(1.52,14) = 22.74$, $p < 0.001$, S2 range: $F(2.25,14) = 32.08$, $p < 0.001$], but the two variables did not interact [FOCUS * REGISTER S1 range: $F(7.8,14) = 1.18$, $p = 0.32$, S2 range: $F(10.45,14) = 1.53$, $p = 0.13$].

The statistical test therefore shows that, while the register differences induced a significant difference in the realization of the rise-fall tone sandhi contour on both syllables, this register effect remained intact under the different focus conditions, and was not maximized or minimized in the presence or absence of focus. For this reason, Figure 4.1 displays the four register combinations on the two syllables pooled over the five focus conditions. The separate effect of focus will be discussed in the next subsection.

As can be seen in Figure 4.1, the effect of register is clearest in the early portions of the syllables, which are higher for syllables with a high register tone than for syllables with a low register tone on both syllables. As explained in section 4.2.1, this effect is a by-product of a co-occurrence constraint, according to which high register tones are realized on syllables with a voiceless onset, and low register tones are realized on syllables with a voiced onset. Consequently, on syllables with a voiced onset, the rise-fall sandhi contour will be realized lower than on syllables with a voiceless onset. The register effect is most visible in the earlier portions of both syllables, whereas it subsides in later portions of the tonal contour.

As shown in the statistical test, the register effect does not interact with the focus conditions. For this reason, the next section will discuss the focus effects averaged over all four register combinations.

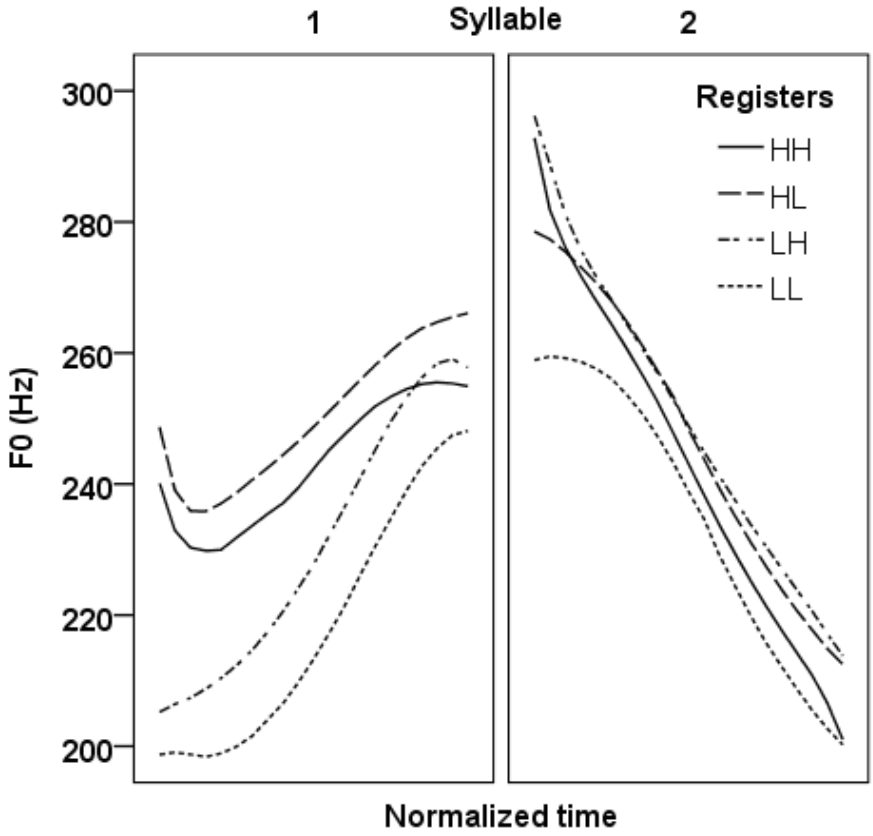


Figure 4.1: Averaged and time-normalized F_0 trajectories, pooled across focus conditions, broken down by syllable (left = syllable 1, right = syllable 2) and register combination.

4.3.1.2 Focus effects on F_0

Pooling over the different register combinations in the stimuli makes it possible to perform both by-subjects (*F*1) and by-items (*F*2) RM ANOVAs for the two syllables separately for the factor FOCUS, with F_0 range as the dependent variable. In both the by-items and the by-subjects analysis, syllable 1 and syllable 2 range showed a main effect of FOCUS [S1 range: $F1(1.22,14) = 25.65$, $p < 0.001$, $F2(2.27,11) = 89.22$, $p < 0.001$, S2 range: $F1(1.93,14) = 39.17$, $p < 0.001$, $F2(2.87,11) = 59.39$, $p < 0.001$].

Pairwise post-hoc tests across the five focus conditions showed that there was a significant difference for both syllables between the three on-target focus conditions (S1, S2, Word) and the two non-target focus conditions (Pre-target, Post-target), in that the F_0 range was expanded in the on-target focus conditions. For neither syllable did the difference between the three on-target focus conditions reach significance: S1, S2 range: S1 = S2 = Word. For two non-target focus conditions, for S1 range the by-items analysis returned a significant difference between pre- and post-target focus, but this difference failed to reach significance in the by-subjects analysis. There was no statistically significant difference between pre- and post-target focus F_0 range on the second syllable in either analysis. The findings are summarized in Table 4.1 and Figure 4.2.

Table 4.1: Summary of the statistical findings for F_0 range by focus condition. Focus conditions: S1= Syllable 1 focus, S2 = Syllable 2 focus, Word = Word focus, Pre = Pre-target focus, Post = Post-target focus.

	Syllable 1 range	**Syllable 2 range**
By-subjects (*F*1)	S1, S2, Word > Pre, Post	S1, S2, Word > Pre, Post
By-items (*F*2)	S1, S2, Word > Post > Pre	S1, S2, Word > Pre, Post

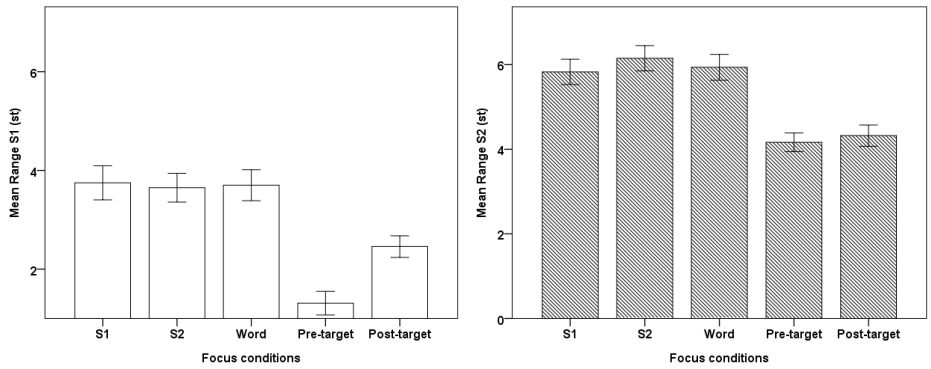


Figure 4.2: Mean absolute pitch change and error bars (±2 SE) for the first (left) and second (right) syllable across five focus conditions. Focus conditions (from left to right): S1 focus, S2 focus, Word focus, Pre-target focus, Post-target focus.

In order to investigate why pre- and post-target focus yield a difference in F_0 range on the first, but not on the second syllable, it is beneficial to look at the actual F_0 curves of the five focus conditions. As can be seen in Figure 4.3, the rising tone on the first syllable in pre-target focus condition starts higher than in all other conditions, and falls a bit before starting to rise later than in the other conditions, thereby reaching a lower F_0 maximum. This is a confound from the fact that in pre-target focus, the syllable preceding the target word, which bears a rising tone, is focused.

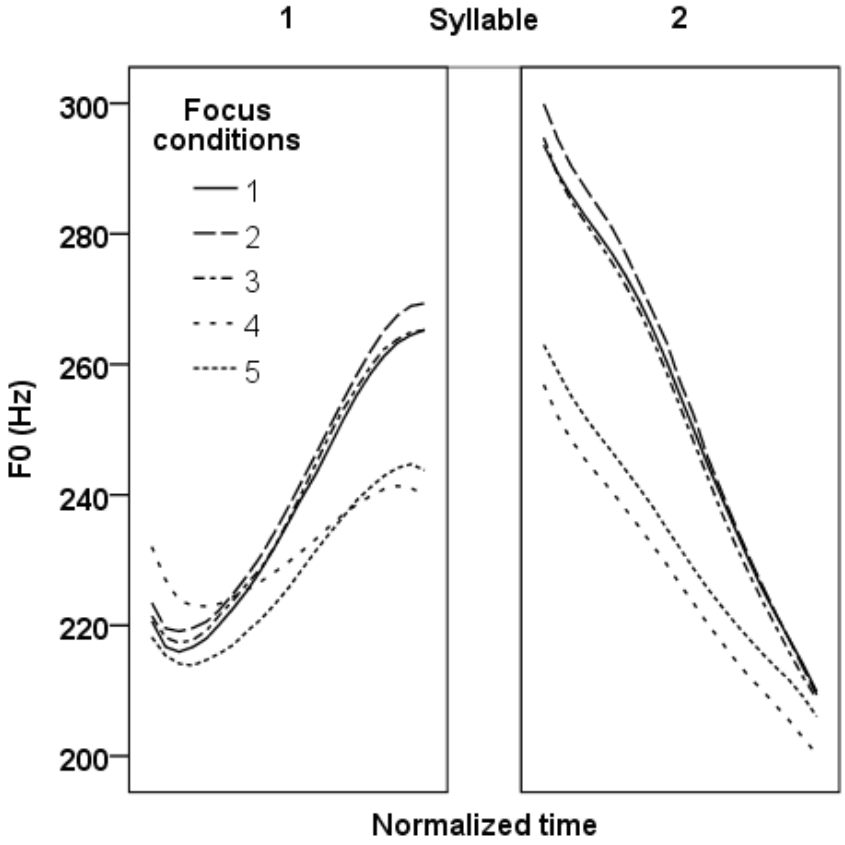


Figure 4.3: Averaged and time-normalized F_0 trajectories across five focus conditions, broken down by syllable (left = syllable 1, right = syllable 2) and focus conditions (1 = S1 focus, 2 = S2 focus, 3 = Word focus, 4 = Pre-target focus, 5 = Post-target focus).

Under focus, this rising tone is realized with a greater F_0 range than in the other four conditions, which is why the tonal contour on the following syllable has to fall first before rising again. After the coarticulatory effect of the previous syllable has worn off, the tonal contour for the pre-target focus is quite similar to that of post-target focus on the second syllable. It can be assumed that in absence of this confound, the pre- and post-target focus condition would also yield comparable effects on the first syllable.

In Figure 4.3, it can furthermore be seen that he F_0 expansion effect of focus seems to be unidirectional. The F_0 range on the focused constituents is expanded mostly upwards, whereas the F_0 minima remain comparable across the five focus conditions. To test whether this impression holds statistically, the F_0 maxima and minima were also compared across the different focus conditions for both syllables. The results are shown in Figure 4.4 below.

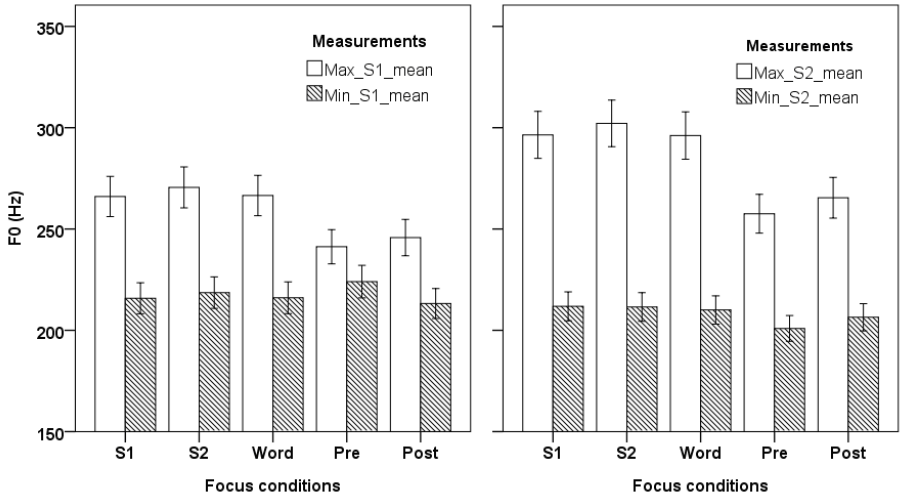


Figure 4.4: Mean F_0 maxima (clear boxes) and minima (patterned boxes) for the first (left) and second (right) syllable across five focus conditions. Focus conditions (from left to right): S1 focus, S2 focus, Word focus, Pre-target focus, Post-target focus. T-bars = ±2 SE.

Two RM ANOVAS, again one by subjects (*F*1) and one by items (*F*2), both yielded main effects for the factor FOCUS on both syllables for all measurements: Syllable 1 F_0 maximum [$F1(1.56,14) = 35.63$, $p < 0.001$, $F2(3.66,11) = 100.03$, $p < 0.001$], Syllable 1 F_0 minimum [$F1(1.41,14) = 4.65$, $p = 0.032$, $F2(2.14,11)$

= 7.69, p < 0.003], Syllable 2 F_0 maximum [*F*1(1.43,14) = 48.07, p < 0.001, *F*2(3.91,11) = 140.75, p < 0.001], and Syllable 2 F_0 minimum [*F*1(2.98,14) = 9.95, p < 0.001, *F*2(2.41,11) = 12.66, p < 0.001].

Post-hoc pairwise comparisons across the five focus conditions, as illustrated in Table 4.2, revealed a similar picture for the F_0 maxima on both syllables as the F_0 range measurements: the F_0 maxima were significantly different between the three on-target focus conditions and the two non-target focus conditions, but there was no significant difference within either group for either analysis. For the F_0 minima, there was no such clear division between the different target conditions, and only some spurious significances surfaced, mostly involving the raised F_0 minimum on the first syllable of the target word due to the preceding rising tone.

Table 4.2: Summary of the statistical findings for F_0 range by focus condition. Focus conditions: S1= Syllable 1 focus, S2 = Syllable 2 focus, Word = Word focus, Pre = Pre-target focus, Post = Post-target focus.

	By-subjects (*F*1)	By-items (*F*2)
S1 F_0 maximum	S1, S2, Word > Pre, Post	S1, S2, Word > Pre, Post
S1 F_0 minimum	S2 > Post	Pre > Post S2 > Post
S2 F_0 maximum	S1, S2, Word > Pre, Post	S1, S2, Word > Pre, Post
S2 F_0 minimum	S1, S2, Word > Pre	S1, S2, Word, Post> Pre

The results show that the significant effect of focus on the F0 range on the two syllables of the target word is brought about more or less exclusively by a raising of the F0 maxima, whereas the F0 minima remain relatively unaffected by the focus conditions, and only show secondary coarticulatory effects. For the F0 minima, there is no systematic way to distinguish all on-target focus conditions from all non-target focus conditions, which shows that the F0 expansion effect of focus in Wenzhou is unidirectional.

4.3.2 Duration effects

For the duration measurements of both syllables, again a by-subjects (*F*1) and by-items (*F*2) RM ANOVA was conducted with FOCUS as dependent variable. Both analyses returned main effects of FOCUS on the duration measurements: Syllable 1 duration [$F1(3.27,14) = 25.49$, $p < 0.001$, $F2(4,11) = 15.75$, $p < 0.001$], and Syllable 2 duration [$F1(2.25,14) = 36.17$, $p < 0.001$, $F2(3.72,11) = 41.2$, $p < 0.001$]. The measurements are illustrated in Figure 4.5 below.

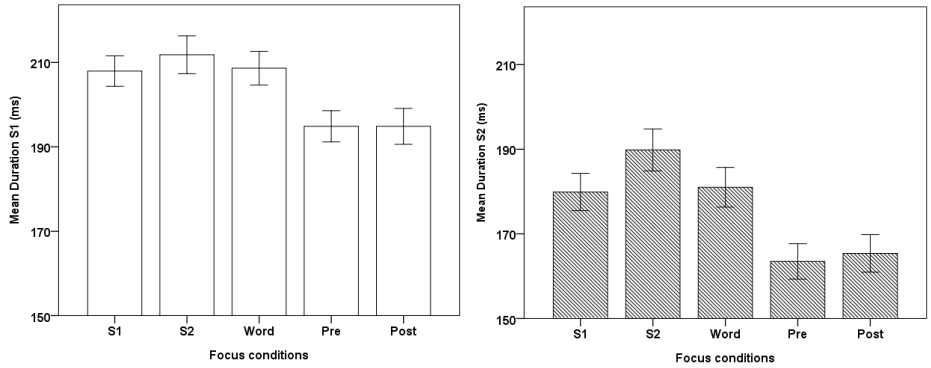


Figure 4.5: Mean duration of syllable 1 (left) and syllable 2 (right) of the target word across five focus conditions. Focus conditions (from left to right): S1 focus, S2 focus, Word focus, Pre-target focus, Post-target focus. T-bars = ±2 SE.

The post-hoc pairwise comparisons across the five FOCUS conditions again showed a significant difference between the three on-target focus conditions (S1, S2, Word) and the two non-target focus conditions (Pre-target, Post-target) on the first syllable, in that its duration was longer under focus. However, the detailed comparisons on the second syllable returned different results for the two analyses. While in the by-items analysis there was no significant difference between the three on-target focus conditions in either measurement (S1 = S2 = Word), the by-subjects analysis showed a significant difference for the length of the second syllable: it was longer under narrow focus (S2 focus) than when focus was on the first syllable (S1 focus).

However, the difference in duration of the second syllable under S2 focus did not reach significance in comparison with the Word focus condition. Pre- and Post-target focus did not return a significant difference in either analysis. The findings are summarized in Table 4.3.

Table 4.3: Summary of the statistical findings for syllable duration by focus condition. Focus conditions: S1= Syllable 1 focus, S2 = Syllable 2 focus, Word = Word focus, Pre = Pre-target focus, Post = Post-target focus.

	Syllable 1 duration	**Syllable 2 duration**
By-subjects (*F*1)	S1, S2, Word > Pre, Post	S2 > S1 S1, S2, Word > Pre, Post
By-items (*F*2)	S1, S2, Word > Pre, Post	S1, S2, Word > Pre, Post

4.4 Discussion

4.4.1 Summary of results

The results of the experiment show that in Wenzhou Chinese as in other Chinese dialects, contrastive focus is marked both by F_0 range expansion and lengthening. However, neither of the two effects allows for a systematic distinction between the three on-target focus conditions, or between the two non-target focus conditions.

From an F_0 perspective, the phonetic effects of focus on the first syllable, the second syllable, or the entire target word within the disyllabic tone sandhi domain of Wenzhou Chinese are similar. This means that, in Wenzhou Chinese unlike in many intonational languages, focus cannot pick out one syllable in a word as the location of focus and mark it with a distinct F_0 movement to distinguish it from other syllables within that word. Rather, the F_0 contour that results from disyllabic tone sandhi remains intact under focus: even when focus undercuts the tone sandhi domain, the phonetic reflex of focus is similar to when the entire word is in focus.

In terms of duration, there is a clear difference between the three on-target focus conditions and the two non-target focus conditions, which manifests itself in a stable lengthening effect on both syllables. Additionally, under S2 focus, the second syllable is lengthened to a greater extent than under S1 focus, while no comparable lengthening effect of S1 focus is visible on the first syllable. However, the duration of the second syllable under S2 focus is not significantly different from the duration of the same syllable under Word focus, and only reaches significance compared to S1 focus in one of the two statistical

tests. This leads to the conclusion that, while there is slightly greater lengthening on the second syllable when it is in narrow focus, the duration effects on neither syllable are big enough to allow distinction between all three on-target focus conditions.

Besides the absolute lengthening effect of focus on both syllables, it is also interesting to look at the relative duration of the two syllables with respect to one another. As can be seen in Figure 4.5, it is the case in all the focus conditions that the duration of the first syllable surpasses that of the second syllable. This is comparable to the edge-effect described in Chen 2006, which states that word-initial syllables tend to be longer than medial syllables (*ceteris paribus*). In the current experiment, the quadrisyllabic phrase containing the disyllabic target word and the disyllabic noun which it modifies, shows a comparable pattern of internal length distribution to the quadrisyllabic words in Standard Chinese described in Chen 2006.

However, the effect of lengthening under focus is different from that described for Standard Chinese, in that the durational distribution within the target word (initial syllable always longer than second syllable) remains intact under all focus conditions. In contrast, in Chen 2006, the duration of the second syllable under S2 focus exceeded that of the first syllable. This finding presents further evidence for the conclusion that the lengthening effect of focus in Wenzhou is distributed over both syllables of a disyllabic word, regardless of the exact position of the focus within that word.

As for the pre-and post-target conditions, they prove indistinguishable, except where external factors affect either part of the measurement domain (cf. Figure 4.3). When looking at the syllable 1 F_0 range, it is smaller in pre-target focus condition than in post-target focus condition, because the (focused) rising tone on the pre-target syllable expands its F_0 range and thereby diminishes the F_0 range of the immediately following syllable. However, when comparing the F_0 maxima and minima independently, it can be seen that only the F_0 minima are affected by this coarticulation, whereas the F_0 maxima only reflect focus condition. Furthermore, the present experiment finds neither a lengthening nor an F_0 effect that distinguishes material in pre- from post-focus position.

4.4.2 Implications of the findings

The current study has shown that the phonetic mechanisms of marking focus on sandhi tones, namely F_0 range and duration expansion, are similar to those that have been reported for lexical tones, a finding that was already discussed for the sandhi tones in Shanghai Chinese in Chen 2009. However, unlike in Shanghai

Chinese, no significant post-focal lowering of F_0 contours was found in the present experiment. This gives reason to conclude that Wenzhou, unlike Standard Chinese and Shanghai Chinese, does not employ post-focal lowering as an additional cue to focus location.

In Wenzhou Chinese, the realization of contrastive focus in the disyllabic word domain, which is the domain of phonological tone change in this language, is influenced by the specific characteristics of this domain. What has been shown is that the focus effects of F_0 range expansion and lengthening are distributed over the entire disyllabic domain, even when only one of the two syllables in this domain is the target of the contrastive focus. This is in contrast to findings for other dialects of Chinese, such as Standard Chinese (Chen 2006) and Hong Kong Cantonese (Gu & Lee 2007b), for which it has been shown that contrastively focused syllables within polysyllabic words can receive greater phonetic marking than other syllables in the same word.

By contrast, in Wenzhou Chinese, it appears that the disyllabic tone sandhi domain cannot be split up by contrastive focus, so that one syllable would receive greater phonetic marking than the other. Rather, both the F_0 contour and the duration distribution of the disyllabic words are similar under focus on the whole word or on one of its parts. It appears that focus can only affect the entire disyllabic tone sandhi domain as a whole, but not break up its internal structure.

This is in contrast to what has been found for intonation languages such as Dutch, where the marking of the exact location of focus can take precedence over the phonetic display of phonological properties such as lexical stress, and for example shift the location of a pitch accent within a word. Via the segmental durations, the speaker still receives cues to locate the lexical stress properly, but the primary cue, namely the pitch accent, is utilized for focus marking. In Wenzhou Chinese on the other hand, it seems that the realization of the tone sandhi F_0 contour mainly serves to convey the cue of disyllabic wordhood (rather than just two syllables that happen to be adjacent), at the expense of marking the exact location of contrastive focus for the listener.

While there is a minimal effect of duration, which is extended on the second syllable when this syllable is narrowly focused, it seems that also the durational marking is mostly uniformly expanded over the entire disyllabic domain under focus. This is similar to what has been found for Dutch (Sluijter & van Heuven 1995), namely that while focus induces a lengthening of the word under focus, the internal durational distribution between the syllables of the word remains mostly intact. In that sense, speakers receive a (durational) clue

for stress in Dutch even under focus, and analogously it can be said that they receive a (durational) clue for the phrase-initial position (i.e. the first syllable being longer than the second), even when focus marking would favor lengthening on the second syllable.

These findings can be interpreted to mean that the speakers conceptualize the entire disyllabic tone sandhi domain as one whole, and that the integrity of the domain is preserved under focus. Rather than individually expanding the F_0 range or duration of one of the syllables, the speakers expand the contour on both syllables upwards to strengthen the tonal realization. This finding also speaks for an interpretation of the tone sandhi process whereby, as soon as two lexical tones come together in a disyllabic compound word, these tones are "replaced" by a tonal contour that is spread over the whole disyllabic word domain. Additional effects such as focus marking can then only affect this contour as a whole, but not break it up into its components any more to emphasize one over the other.

At the same time, the findings speak for a view of focus marking that has to allow a greater room for phonological processes or prosodic constituents. While it seems that the straightforward effects of focus on F_0 and duration lend themselves to an analysis that sees focus as something that is phonetically implemented, such a conclusion would be at odds with the results of the current experiment. If focus were just a phonetic effect that gets added to the finished derivation, it would be counterintuitive to expect that it should pay attention to the lexical integrity of compound words in one dialect (Wenzhou), but not in others (Standard Chinese).

Rather, it seems that the current findings lend themselves to a more indirect view of focus (see also Chen 2009 and Chen & Gussenhoven 2008), which sees the effects of focus as comparable to the strengthening effect of prosodic prominence within prosodic constituents. Under such a view, the F_0 range and duration expansion observed under focus is then a consequence of a more abstract, phonological "strengthening" effect brought about by focus, rather than the manifestation of a phonetic focus effect itself. Such a more indirect effect of focus, which is mediated by prosodic structure specifications in the respective language, can also help to explain recent findings for multiple focus in Standard Chinese (Kabagema-Bilan et al. 2011).

For the present experiment, it is clear that prosodic structure plays an important role in the implementation of focus, namely by constraining how narrowly focus can be marked. The results here suggest that the tone sandhi domain, within which changes to the F_0 contour of syllables within disyllabic

words take place, is also the domain that limits the distribution of focus marking, at least when it comes to F_0 effects. When computing the tonal contours of disyllabic compounds, Wenzhou speakers have to take the tonal information on both syllables into account, and see the entire disyllabic word as one whole. This holistic perspective is reflected in focus marking by F_0 expansion, which applies uniformly across the two syllables within the domain, no matter whether the actual focus domain is the whole word or either of the two syllables.

It is interesting to compare the findings of the current experiment to the (brief) description of focus effects in Wenzhou in Chen 2000. In this book chapter, which is based on the impressionistic description of recordings from a middle-aged speaker in the 1980s, it is not only predicted that focus should be able to single out individual syllables in the disyllabic compound domain, but also that it should be able to break up the disyllabic prosodic domain at all, and interrupt the phonological process of tone sandhi within it (so that each syllable would be realized individually with its lexical tone).

Such a phonological view of focus, in which the focus can directly manipulate the presence/absence of prosodic boundaries and the phonological processes that are connected to them, has been showed to be inaccurate on other grounds before (Chen 2004). For the young speakers recorded in the current experiment, it seems that the limitations of focus marking are yet one step further ahead: not only does the tone sandhi contour remain intact in the presence of a focus that singles out an individual syllable within it, but even the phonetic implementation of the focus effect appears to be mediated by the prosodic tone sandhi domain as a whole.

Therefore, the need for a more "phonological" view of focus effects in Chinese should not be taken to mean that focus should be able to modify the prosodic structure as it is mapped from syntax. Rather, focus appears to be sensitive to the prosodic structure and its specifications, but unable to change its direct components. Rather, this prosodic structure seems to limit the extent to which focus can modify the tonal information, while still ensuring that crucial parts of the information (such as "wordhood" in the current experiment) remain intact.

4.5 Conclusion

This study investigated the distribution of narrow focus marking within the disyllabic tone sandhi domain of Wenzhou Chinese. An experiment looked at the influence of sub-word focus on the implementation of the rise-fall sandhi

contour in disyllabic words. Recordings from 15 young Wenzhou speakers were analyzed, in which they read out question-answer pairs which induced contrastive focus on either or both syllables of the disyllabic target word. Additionally, focus on the pre- and post-target word was tested for comparison.

Analysis of the F0 curves of the recordings shows that, regardless of the exact location of focus within the disyllabic tone sandhi domain, the tonal contours on both syllables are modified in a similar manner, compared to the control conditions (i.e. pre- and post-target focus). Lengthening likewise targets both syllables under focus on the whole word, and there is no significant difference between word focus and focus on either syllable. The second syllable shows slightly greater lengthening under syllable 2 focus than under syllable 1 focus, but this duration difference is not enough to reliably distinguish all three focus conditions from one another.

These findings suggest that focus affects the realization of the entire tone sandhi contour, even when only one of the syllables is contrasted. Sandhi tones are derived within the disyllabic domain, and likewise the F0 effects of focus are spread out over the entire domain. For duration, the only difference between the conditions is on the second syllable when it is narrowly focused, but the small size of the effect suggests that the entire disyllabic tone sandhi domain can still be considered as the location of the prosodic implementation of focus effects. In short, the disyllabic tone sandhi domain limits the distribution of focus effects in Wenzhou Chinese, which argues for a non-direct mapping of focus domain with the prosodic marking of focus.

Chapter 5

Tonal coarticulation as prosodic marker in Wenzhou Chinese

5.1 Introduction

5.1.1 Tonal coarticulation in Chinese

The implementation of lexical tones in tone languages is influenced by neighboring tones, just as segments coarticulate with other segments. A number of instrumental studies across different languages have shown that neighboring tones influence each other in a way that affects the realization of both tones, but have also pointed out cross-linguistic differences in the exact details of these coarticulatory influences.

For example, the realization of lexical tones in Thai (Gandour et al. 1994; Potisuk et al. 1997), Mandarin Chinese (Xu 1997), and Taiwan Min (Wang 2002) in naturalistic speaking conditions has been shown to be asymmetrically influenced by neighboring tones, with carryover coarticulation exerting a greater influence than anticipatory coarticulation. When Mandarin speakers were instructed to keep their speaking manner constant and not implement the natural stress difference between syllables, however, it could be shown that the two effects were similar in strength (Shen 1990b). For naturalistic speech in Malaysian Southern Min, it has also been argued that anticipatory and carry-over coarticulation are comparable in magnitude (Chang & Hsieh 2012).

In all cases, it was found that the coarticulation effect was mainly localized in the part of the syllable that was adjacent to the influencing context, and that the influence of the neighboring syllable's tone decreased with greater distance from that syllable. However, the exact nature of the coarticulation has also been shown to differ between languages. For Thai and Mandarin Chinese for example, it was argued that carryover coarticulation is assimilatory in nature, such that e.g. a high tonal offset before a rising tone on the following syllable raises the onset of that rise. In contrast, anticipatory coarticulation was found to be primarily dissimilatory, such that speakers magnified the differences between two successive tonal targets across syllable boundaries (Gandour et al. 1994; Xu 1997). In contrast, Peng (1997) for Taiwan Min, Han & Kim (1974) for

Vietnamese, and Wang, H. S. (2002) for Malaysian Southern Min showed that the manner of tonal coarticulation (assimilation/dissimilation) varied more between the individual tones than between the direction of coarticulation (anticipatory/carry-over).

Furthermore, languages also vary with respect to the aspect of the tonal production that is influenced by coarticulation. For Thai, it has been argued that coarticulation mainly affects the height aspect of tonal implementation (Gandour et al. 1994), whereas for Mandarin, slope seems to be the variable that is affected most (as in Xu 1994, but see Shen 1990b for different findings). For Vietnamese, both parameters seem to be affected by coarticulation (Han & Kim 1974). In sum, while neighboring tones affect the realization of tonal contours in all tone languages that have been studied in detail, the languages may vary in the exact type, extent, and direction of the influences from one tone onto another.

5.1.2 Contextual influence on coarticulation

For segmental coarticulation, a number of studies have shown that its extend depends on different contextual factors. For example, prosodic structure has been shown to influence the magnitude of coarticulation in several respects. In contexts where two segments are separated by prosodic boundaries of different levels (prosodic word, prosodic phrase, intonational phrase, utterance), it has been shown that they coarticulate less with each other if they are separated by a higher-level prosodic boundary, compared to a lower-level prosodic boundary (Byrd & Saltzman 1998; Cho 2004, 2006; Fougeron & Keating 1997; Jun 1998).

At the same time, it has also been shown that the strength of a syllable influences the magnitude with which it coarticulates with neighboring segments. Specifically, a greater coarticulation effect has been shown for unaccented compared to accented syllables (Cho 2004), and for unstressed compared to stressed syllables (Cabré & Prieto 2005; de Jong et al. 1993). While prosodic strength of a syllable and the magnitude of a neighboring prosodic boundary may be connected, it has also been shown that the two effects are to some extent independent of each other (Cho & Keating 2009). In that sense, segmental coarticulation serves as a marker for both prosodic boundary level and positional prosodic strength (Cho 2011).

Most research on the phonetic correlates of prosodic structure has been concerned with intonation languages like English or French. Only a few studies have looked at segmental effects in tone languages (Cao & Zheng 2006; Hayashi et al. 1999; Hsu & Jun 1998; Pan 2007a; Zheng et al. 2006) or included them as part of a cross-linguistic comparisons (Keating et al. 2003). All the above

mentioned studies only included two or three speakers each, and some reported contradictory results. For example, of the two Taiwanese speakers investigated in Hayashi et al. 1999 and Keating et al. 2003, only one distinguished more than two levels of prosodic structure in the strength of the articulation of post-boundary consonants. On the other hand, Pan (2007a) reports that for her three Taiwanese speakers, the coarticulation of pre-boundary segments varied with the strength of the intervening boundary, with higher boundaries inducing less nasal coarticulation.

Segmental coarticulation was also found to be reduced across higher prosodic boundaries in Mandarin Chinese by Zheng and colleagues (2006), compared to lower prosodic boundaries. However, the question whether prosodic boundaries also affect tonal coarticulation has only been asked once by Pan & Tai (2006) in a small-scale study with three speakers. They report no statistics, but observe that the F_0 range of falling tones is greater when this tone precedes or follows an IP boundary than preceding or following a lower prosodic boundary.

The current chapter proposes to investigate prosodic structure by looking at the realization of the lexical tones in Wenzhou Chinese. The idea is as follows: if prosodic boundaries and prosodically strong positions (prosodic heads) can induce a strengthening and coarticulatory resistance for segments, the same should happen to lexical tones in tone languages. Consequently, if prosodic levels are the primary determiner of tonal coarticulation, lexical tones at the edges of higher prosodic constituents should coarticulate less with neighboring tones than lexical tones at the edges of lower prosodic constituents. If prosodic strength is the main factor to influence tonal coarticulation, tones in prosodically strong positions should show greater resistance to coarticulation, and be implemented in a more independent way compared to lexical tones in prosodically weak positions.

5.1.3 Focus effects on tonal coarticulation

Recent years have seen a number of studies investigating the effects of focus on the implementation of lexical tones. Across these studies, “focus” as a term is used both for contexts in which a speaker corrects or contrasts a part of an utterance (sometimes called ‘contrastive focus’), and for contexts in which a speaker answers a question containing a wh-expression (sometimes called ‘information focus’), and in both types, the constituent corresponding to the wh-expression or the corrected part of the context sentence is considered to be focused.

To mark these focused constituents, it has been found that, outside of lengthening, F_0 modification is applied to focused syllables and words in Mandarin or Standard Chinese (Wang & Xu 2006; Xu 1999), Cantonese (Gu & Lee 2007b), Shanghai Chinese (Chen 2009), and Taiwan Mandarin (Chen et al. 2009; Xu et al. 2012). In these dialects, the tones on focused constituents are implemented within an expanded F_0 range under focus, such that high targets were realized higher and low targets were realized lower than in control contexts.

However, taking a closer look at the mechanisms of F_0 expansion of tonal contours under focus in Chinese, Chen & Gussenhoven (2008) for Standard Chinese and Chen (2009) for Shanghai Chinese have argued that taken together, the F_0 adjustments lead to an enhanced implementation of the tonal contours that cannot just be reduced to F_0 expansion. For example, the rising/falling tonal trajectories in Chen & Gussenhoven 2008 were implemented with magnified and more distinct movement gestures under focus, and exhibited less coarticulatory influence from neighboring tones.

Similarly, Chen (2010) argues that tonal articulation in post-focal position, which has been assumed to be lowered or compressed compared to focused positions (Jin 1996; Xu 1999, 2005), is better accounted for as hypoarticulation. In a mirror image to the strengthening of tonal implementation under focus, tones in post-focal positions are better understood as being weakly implemented, similar to tones in prosodically weak elements (Chen & Xu 2006). A by-product of this weak implementation is a greater susceptibility to coarticulatory influence from neighboring tones (see also Chen & Gussenhoven 2008).

In sum, there is evidence that speaks for a strengthening of tonal implementation under focus, which also reduces the amount of coarticulatory influence of neighboring tonal contours onto the strengthened syllable. The current chapter will investigate whether this strengthening under focus is similar to the strengthening of segments in prosodically strong positions, and whether both types of strengthening exert a comparable influence onto the magnitude of tonal coarticulation.

5.1.4 Current experiments and hypotheses

The current study was designed to answer two connected research questions: (i) which contextual factors affect tonal coarticulation in Wenzhou Chinese, and (ii) how do these factors interact with the strengthening of tonal implementation induced by focus? In accordance with the findings in Xu 1994, 1999 for Standard Chinese, the amount of coarticulatory influence was measured as a

change in tonal slope in the implementation of the tones on the target syllables. In the following, the first experiment is described, which tests the predictions on prosodic structure. The predictions on the influence of focus will be tested in a second experiment, presented in section 5.5.

Since it has been shown in the investigations on segmental coarticulation that prosodic boundaries (e.g. prosodic word vs. prosodic phrase boundary) and prosodic strength (such as stress) exert different types of influence on the coarticulation of adjacent segments (see section 5.1.2), both elements of prosody are taken into consideration here. Specifically, the test phrases were designed in a way that they would show different results if tonal coarticulation was mainly boundary-dependent, or if it was mainly dependent on prosodic strength.

In order to tease apart the prosodic boundary from prosodic headedness effects, two types of morphosyntactic structures involving a syntactic VP were investigated: adverb-verb structures and verb-object structures. In most syntax-to-prosody mapping algorithms, a distinction is made between arguments and adjuncts of a VP. While arguments can form a prosodic domain together with their heads, adjuncts are mapped onto a separate prosodic domain (Gussenhoven 1992; Samek-Lodovici 2005; Truckenbrodt 1995, 1999). [16] Under this assumption, the two structures outlined above would result in two different prosodic outputs. The adverb-verb structures would be separated by a prosodic phrase boundary, since the adverb is adjoined to the VP, while the verb-object structure would be mapped into one prosodic phrase together.[17]

[16] In impressionistic accounts of syntactically conditioned tone change in Chinese dialects, it has been argued that for the purpose of prosodic phrasing, VP-adverbs have to be treated differently from sentential adverbs to account for the differences in perceived tonal realization (Chen 1987, 2000; Lin 1994; Soh 2001). In these studies, observations about tonal realization were taken as evidence, and the syntactic analyses were based on these observations. In contrast, the current study presupposes an unambiguous analysis of the syntactic structures, and tests the influence of the derived prosodic structures in an instrumental and quantified way.

[17] For the purpose of syntax-to-prosody mapping, it is irrelevant whether a preverbal adverb is analyzed as adjunct to the VP, or to a higher functional projection like IP or TP, since both structural configuations result in a similar prosodic mapping. Likewise, it does not matter for prosody whether adverbs are analyzed as adjuncts, or as specifiers in empty-headed functional projections. Both analyses would have the same consequence for syntax-prosody mapping, namely the pre-verbal adverb being positioned outside of the core VP. See for example Alexiadou 1997; Cinque 2004; É. Kiss 2009 for recent discussion of the syntactic analysis of adverbs.

At the same time, the two structures also differ in the prosodic positional properties of the two constituents involved, namely in the position of the prosodic heads. In intonation languages, prosodic headedness is usually manifested in stress or nuclear accents on prosodically strong positions (see e.g. Fougeron 1999 for an overview). In tone languages such as Chinese, prosodic heads are commonly associated with the preservation of tonal features (Chen 2000; Yip 1999). This means that lexical tones in prosodically strong positions tend to be preserved and articulated clearly, whereas tones on non-head syllables are prone to tone change induced by neighboring tones (Yip 1999).

For the two prosodic structures tested here, the difference in morphosyntactic configuration results in a different distribution of prosodic headedness. For the adverb-verb structures, which are assumed to map onto two prosodic phrases, each of the phrases has its own prosodic head (or prominent position), in line with the common assumption of a one-to-one relation between prosodic constituents and prosodic heads (Hayes 1995). In the case of the verb-object structures, which are phrased in one prosodic phrase together, the common assumption of "nonhead prominence" or "Non-Head Stress" (Duanmu 1995, 2005, 2007, 2012) determines that the constituent that is not the syntactic head is assigned prosodic prominence and thereby attains prosodic head status (Chen 2000; Cinque 1993).

This means that the two structures tested in the current experiment not only differ along the dimension of prosodic boundary between the target word and the surrounding tonal context (prosodic word vs. prosodic phrase boundary), but also along the dimension of prosodic headedness. While the adverb-verb structure maps onto two prosodic phrases, each of which gets its own prosodic head, in the verb-object structure only the object is promoted to phrasal head status, as illustrated in Figure 5.1 (prosodic headedness marked by asterisk).

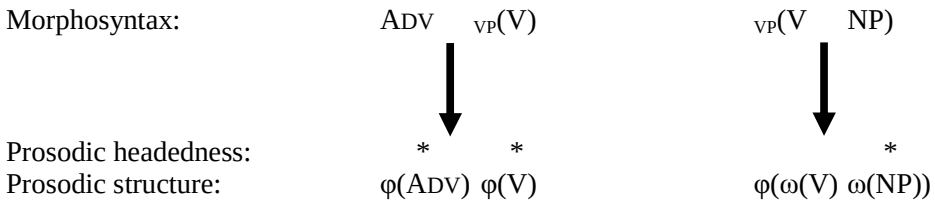
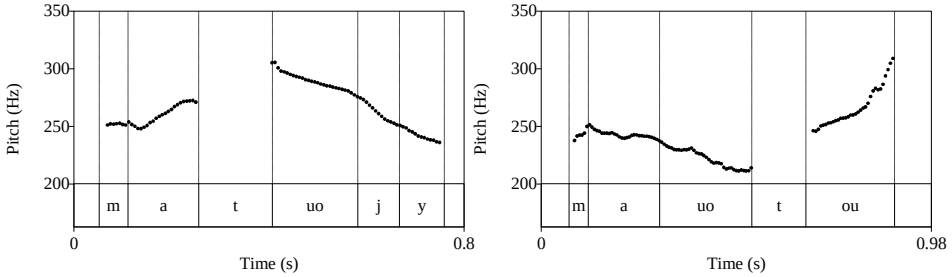


Figure 5.1: Schematic illustration of the assumed syntax-to-prosody mapping and the resulting difference in prosodic headedness.

In order to test the influence of prosodic boundaries and prosodic headedness on tonal coarticulation independently, the position of the target monosyllable was varied within the above structural configurations (target words shown in bold). For the adverb-verb structures, tonal coarticulation will be tested for monosyllabic target words that function as adverbs ((**ADV**)(V-V)) and as verbs ((ADV-ADV)(**V**)), and likewise for the verb-object structures, it will be tested both for monosyllabic target words that function as verbs (**V**(O-O)) and as objects ((V-V)**O**). Note that in all cases, the constituent order within the phrases remains unchanged.

In order to keep variations in sentence lengths and possible syntactic parsings as small as possible, the size of the stimuli was limited to three syllables. All trisyllabic stimuli consisted of a combination of a monosyllabic and a disyllabic lexical target word, whereby the monosyllabic target word could appear on either leftmost or rightmost within the stimulus phrase. In Wenzhou, disyllabic lexical words regularly undergo phonological tone change, whereby the tonal target of both syllables changes from the lexical tone to a specific tonal contour (disyllabic tone sandhi, see Chapter 2 of this thesis for a detailed description).



(1) a. Compatible context

(V	(N-	N))
ma	*tuɔ*	*jø*
24	(42.	31)
buy	dumpling.soup	

'to buy dumpling soup'

b. Conflicting context

(V	(N-	N)
ma	*ʔuo*	*tou*
24	(22.	35)
buy	soy.beans	

'to buy soy beans'

Figure 5.2: Example realization of the same experimental target syllable (initial) by the same speaker in compatible (left) and conflicting (right) context.

In order to test the influence of the context on tonal realization, the target words where elicited both in conflicting and in compatible contexts (see Xu 1994). Figure 5.2 illustrates an example, in which the same target word (*ma* with a rising tone) was followed by two different disyllabic context words. In the left example, the rising target word is followed by a high falling tone on the first syllable of the disyllabic context word, which represents a compatible context. In the right example, the rising target word is followed by a low level tone on the first syllable of the disyllabic context word, which represents a conflicting context.

Including both conflicting and compatible tonal contexts in the experiment serves to test the general prediction that the differences in tonal realization are connected to the adjacent tonal context. The expectation is that, in general, tones should be realized with steeper tonal contours in compatible context, because the adjacent tonal targets can be reached by the speakers without adjusting the tone realization on the target word. In comparison, tonal realization in conflicting contexts should induce adjustment in the implementation of the tones. Therefore, a comparison between the implementation in conflicting and compatible context can give a first indication of tonal coarticulation.

Furthermore, testing trisyllabic phrases with the two morphosyntactic structures outlined above, different predictions emerge with respect to prosodic boundary vs. prosodic head effects. For the first experiment, three different outcomes are conceivable.

(i) No tonal coarticulation difference between the two prosodic structures

Leftmost target words	Rightmost target words
(**ADV**)(VV) = ((**V**)(O-O))	(ADV-ADV)(**V**) = ((V-V)(**O**)

Possible interpretations:

- Tonal coarticulation in Wenzhou Chinese is not dependent on prosodic structure
- Or: Tonal coarticulation in Wenzhou Chinese is dependent on prosodic structure, but the two morphosyntactic structures that were tested map onto identical prosodic structures

(ii) More tonal coarticulation in verb-object structures than in adverb-verb structures on both sides

Leftmost target words	Rightmost target words
(**ADV**)(VV) < ((**V**)(O-O))	(ADV-ADV)(**V**) < ((V-V)(**O**)

Possible interpretations:

- Tonal coarticulation in Wenzhou Chinese is dependent on prosodic boundary strength

(iii) More coarticulatory adjustment in verb-object than in adverb-verb structures in the leftmost target words, but no coarticulatory adjustment difference between the two structures in the rightmost target words

Leftmost target words	Rightmost target words
* * * (**ADV**)(VV) < ((**V**)(O-O))	* * * (ADV-ADV)(**V**) = ((V-V)(**O**)

Possible interpretations:

- Coarticulatory adjustment is dependent on prosodic headedness

A comparison between the tonal implementation of the target words will show which of the predictions above accounts best for the experimental results.

5.2 Method

5.2.1 Stimuli

In order to test the hypotheses made in section 5.1.4, trisyllabic adverb-verb and verb-object structures were designed such that the monosyllabic target words appeared either in leftmost or in rightmost position. In order to control for the context-dependency of coarticulation, all structures were tested in both conflicting and in compatible context. For this purpose, the target words were coupled with disyllabic compounds which carried tone sandhi contours that started or ended with a low or high tonal target.

The monosyllabic target words carried either of the four lexical contour tones of Wenzhou Chinese: low rising, high rising, low falling, and high falling tone. The experiment was limited to contour tones under the assumption that the coarticulatory adjustment effects on slopes would be most clearly quantifiable for these tones, on basis of the steepness of their tonal trajectories. Examples are given in (2) and (3) (target word = bold, tones in Chao numbers).

(2) Conflicting context

	Structure	Hanzi	Wenzhou	Translation
a.	[ADV[V-V]	都喜欢	***ɦu42*** *(sz̩42.cø31)*	‘like everything’
b.	[V[N-N]	喝汤圆	***ha42*** *(t^{h}uɔ42.jø31)*	‘drink dumpling soup’
c.	[[ADV-ADV]V]	赶紧剁	*(kø44.tɕaŋ22)* ***tou42***	‘chop hurriedly’
d.	[[V-V]N]	贩卖报	*(fa44.ma22)* ***pɜ42***	‘sell the report’

(3) Compatible context

	Structure	Hanzi	Wenzhou	Translation
a.	[ADV[V-V]	必学习	***pi42*** *(ɦu24.zai31)*	‘certainly learn’
b.	[V[N-N]	剁猪肉	***tou42*** *(tsɛi35.ɲou31)*	‘chop pork’
c.	[[ADV-ADV]V]	干脆喝	*(kø22.ts^{h}ai33)* ***hɜ42***	‘simply drink’
d.	[[V-V]N]	接收布	*(hu22.lou33)* ***pɜ42***	‘receive a report’

For each of the 32 combinations of factor levels (two prosodic structures*two contexts*two positions for the target word*four lexical tones), two lexically different examples were recorded, which brings the total stimulus count to 64 tokens per speaker. One example was later excluded, because the speakers realized it with a different tonal target than expected, which leaves 63 tokens per speaker for analysis. In addition to the 64 target tokens, the speakers read 80 trisyllabic filler tokens with varying morphosyntactic composition, so that every speaker saw 144 phrases per recording round.

All examples in the experiment were checked by a native speaker of Wenzhou Chinese and three more native Chinese speakers with different dialectal backgrounds for naturalness and grammaticality. The Wenzhou speaker was in the same age group as the participants of the experiment, but did not take part in the recordings himself. Furthermore, the selected phrases were screened for their segmental composition to minimize difficulties in the later analysis process. A full list of stimuli can be found in appendix 5.1 at the end of this thesis.

5.2.2 Speakers

A total of 19 speakers (five males, 14 females) were recorded for this experiment. They all were born and raised in the inner-city Lucheng district of Wenzhou, and were of similar age (mean age = 23.7, SD = 3.0). None of them reported to have lived outside of Wenzhou for a significant amount of time within the last 5 years, and all of them considered themselves fluent speakers of the Wenzhou dialect. All of them were also fluent in Standard Chinese, but had no difficulty reading out aloud Chinese characters in their dialect. None reported any hearing or speech impediments.

Due to a technical error which sometimes cut off the recording of the stimulus before it was completely uttered, the recordings of six speakers had to be excluded because they showed too many gaps per condition. Data of the remaining 13 speakers was analyzed (four male, nine female; mean age = 23.0, SD = 2.8). Four of these speakers read the list of stimuli once, and the other nine speakers read the list twice.

5.2.3 Experimental procedure

Speakers were recorded in individual sessions in a sound-proofed recording studio in the TV and radio station in Wenzhou, and received a small payment for their participation. Each speaker was seated in front of a 13" monitor and given a Sennheiser pc130 headset. The experimenter ensured that the microphone of the headset was placed approximately 3 cm from the corner of the mouth of the subject. Via an external digitizer (UA-G1), the sound was recorded directly on the laptop (Acer Aspire 1810TZ) on which the stimuli were displayed to the speaker.

The speakers were first informed about the recording procedure. They were told that they were supposed to read out phrases and sentences presented on the screen using Wenzhou dialect, in a natural and clear fashion. If they were unsure how to pronounce a word or phrase, they could skip to the next item, and if they felt they had made a mistake, they could go back and repeat the recording of the previous item. They were told that they could interrupt or abort the recording at any point.

Before the actual recording, all speakers completed a practice series with 8 trisyllabic phrases that were not part of the actual experiment. This was done in order to familiarize the speakers with the self-managed recording procedure, during which they had to press a button to initiate the recording of

sound.[18] After completing the practice items, the speakers were asked to indicate whether they understood the recording procedure and were ready to start the actual experiment. Upon confirmation, the experimenter started the actual experiment.

5.2.4 Data analysis

Before data analysis, all recordings were screened for tonal correctness. If a speaker produced a tonal contour on the disyllabic compound that was different from the expected realization in a way that the context was no longer conflicting/compatible, the token was excluded from further analysis. Likewise, all recordings that were incomplete or produced hesitantly with an audible pause within the phrase were excluded. If a speakers who read the list twice made a mistake on one of the recordings, this token was excluded from the analysis. If both recordings were correct, the average of the two recordings was computed and used for the analysis.

Since the target phrases were recorded in isolation, some of the recordings showed utterance-final breathiness and/or creakiness, which obscured parts of the F_0 contour on the final syllable. This concerned a total of 108 recordings, which were marked and subsequently excluded from the F_0 measurements and analysis. A total of 962 target tokens was retained for analysis.

Within the rhyme of the target syllable, F_0 values were sampled at 20 equidistant intervals with the help of the automated F_0 tracking algorithm in PRAAT (Boersma & Weenink 2001). Before data extraction, all F_0 contours were checked for tracking errors such as octave jumps, and these errors were manually corrected (26 instances). Additionally, the contours were smoothed before extraction, using the smoothing function at a 10 Hz bandwidth in PRAAT. A script performed the automated extraction of the duration information and F_0 measurements.[19]

5.3 Results

As described in section 5.1.4, three factors were expected to influence the amount of tonal coarticulation between the tones on the monosyllabic word and

[18] The script used for presenting and recording the stimuli was written by Jos Pacilly, and slightly modified by the author.
[19] The script used for segmenting and measuring the files was written by Jos Pacilly.

the disyllabic compound: (i) the position of the target word (leftmost/rightmost), (ii) the type of context (conflicting/compatible), and (iii) the prosodic structure of the stimulus phrase (ADV-V and V-O). The effects of these three factors will be presented in more detail in the following.

5.3.1 Position

Figure 5.3 illustrates the pitch trajectories in the different prosodic structures and contexts, split by tones and position of target word. In order to be able to average the F_0 values over speakers, all raw F_0 values were converted into semitones and scaled to the individual speaker's pitch range. The speaker's pitch range was set to the averaged values of the turning point in the low rising tone (baseline) and the fall onset in the high falling tone (topline) in rightmost position in compatible context. Figure 5.3 expresses the speaker pitch range on a scale between 0 and 100.

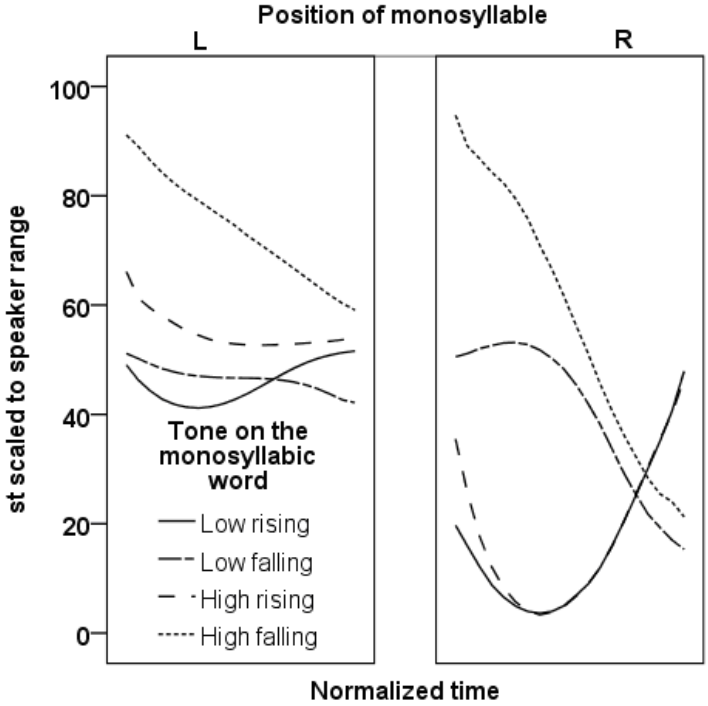


Figure 5.3: Averaged and time-normalized pitch contours of the four investigated tones, broken down by position (L = leftmost, R = rightmost).

As can be seen in Figure 5.3, the realizations of the tonal contours in leftmost and in rightmost position the trisyllabic phrases differ in the amount of the speakers' pitch ranges that they cover. More specifically, the rightmost tones cover a much wider portion of the speakers' pitch ranges than the leftmost tones, when averaged over the different contexts and prosodic structures.

5.3.2 Context

As discussed in section 5.1.2, tonal coarticulation has been found to be context-dependent. Tones coarticulate much more in conflicting contexts than in compatible contexts. In order to illustrate the effect of context, the following graphs displays the four lexical tones in conflicting (top) and compatible (bottom) context in both positions, averaged over the two prosodic structures.

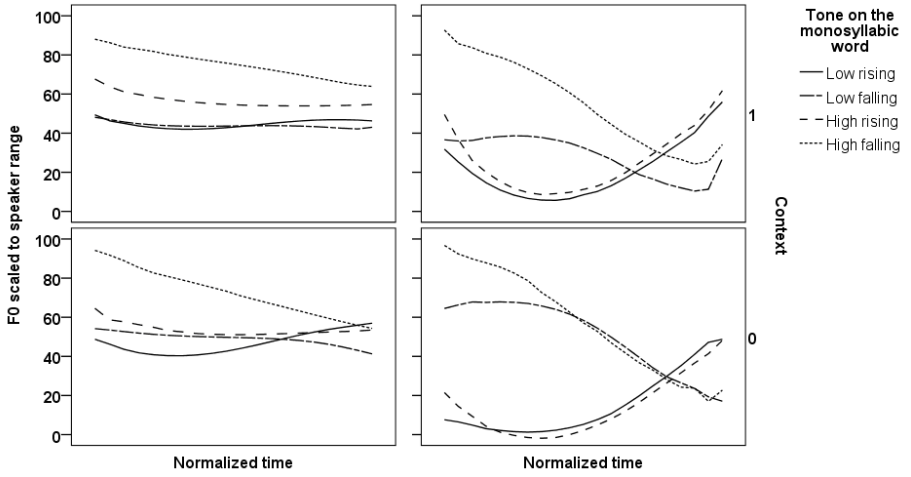


Figure 5.4: Frequency and time-normalized averaged pitch contours of the four investigated tones in leftmost/rightmost position, broken down across contexts.

As can be seen in Figure 5.4, the effect of coarticulation causes minor differences across the entire tonal trajectories, but is most clearly visible in the portion of the tone that is immediately adjacent to the context, i.e. the right edge for the leftmost tones and the left edge for the rightmost tones. In both positions, the context has consequences for the steepness of the tonal realizations. The two rising tones are largely deprived of their final rise in leftmost position in conflicting context compared to compatible context, making their tonal trajectories appear almost flat. In rightmost position, the rising tones start higher

in conflicting compared to compatible context, which also flattens the overall slope. The falling tones both fall less steeply in leftmost position in conflicting compared to compatible context, and they start lower in rightmost position.

5.3.3 Prosodic structure

The previous graphs have indicated that the effect of coarticulation in trisyllabic phrases in Wenzhou is mainly local, and subsides with increasing distance from that tonal context. The investigation in this section will therefore concentrate on the respective halves of the tonal trajectories that are adjacent to the tonal context (i.e. the left half for rightmost targets, and the right half for leftmost targets).

In order to make the differences in the tonal trajectories more clearly visible, the following graphs will display linear approximations to the slopes of the respective halves of the tonal trajectories, rather than the trajectories themselves. In this way, it can be seen more easily whether the two structures induce a difference in tonal implementation. The slope values were computed by dividing the difference in F_0 between the first and last measurement of the trajectory part by half of the duration of the target syllable rhyme. For targets in which either the first or the last measurement were missing (for example because of creakiness or breathiness in the signal), the second or the last-but-one measurement were used to compute the slope value (with accordingly adjusted duration values). Targets which had more measurements missing were not included in the slope analysis.

5.3.3.1 Leftmost targets

In Figure 5.5, the slope values are displayed for the tones in leftmost position ((**ADV**)(V-V) with full lines and ((**V**)(O-O)) with dashed lines). Since coarticulation was mainly found in conflicting contexts, only the conflicting context values are shown.

It can be seen that all four tones display the same tendency: the tonal direction is more preserved in the adverb-verb structures (solid line) than in the verb-object structures (dashed line). For the two rising tones, the tonal trajectories in verb-object structures are almost horizontal (low rising) or even slightly falling (high rising), while the trajectories in the adverb-verb structures have the rising tonal direction preserved. For the falling tones, the falling trajectories are more steeply falling in the adverb-verb structures, and flatter or even rising in the verb-object structures.

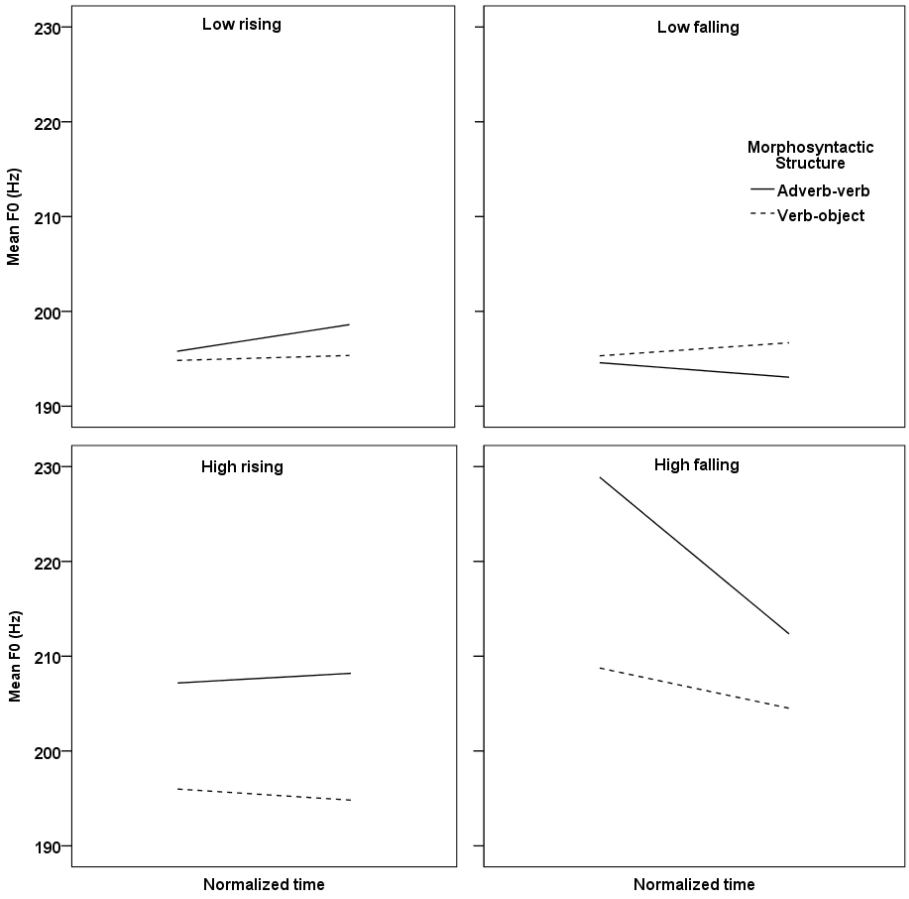


Figure 5.5: Linear representations of the tonal slopes (rightmost half) in leftmost position in conflicting context, broken down by tone on the target syllable and structure (dashed lines = verb-object, solid lines = adverb-verb).

In sum, it can be said that across all four tones, the same tendency is visible: the tonal trajectories are steeper and true to the original tonal direction in adverb-verb structures, but flattened or even slightly reversed in direction for the verb-object structures. In order to statistically test the validity of this observation, the slopes were first "adjusted" (in the sense of Xu 1994) and then pooled over all four tones. For the adjustment, the slope values for the falling tones were multiplied by (-1), so that for all four tones, a positive slope value would

represent a preservation of tonal direction, and a negative slope value would represent a reversal of tonal direction.

Figure 5.6 graphically represents the pooled slope values, split by context (left = compatible context, right = conflicting context) and prosodic structure (light bars = ((**ADV**)(V-V), shaded bars = ((**V**)(O-O))) for the tones on target words in leftmost position.

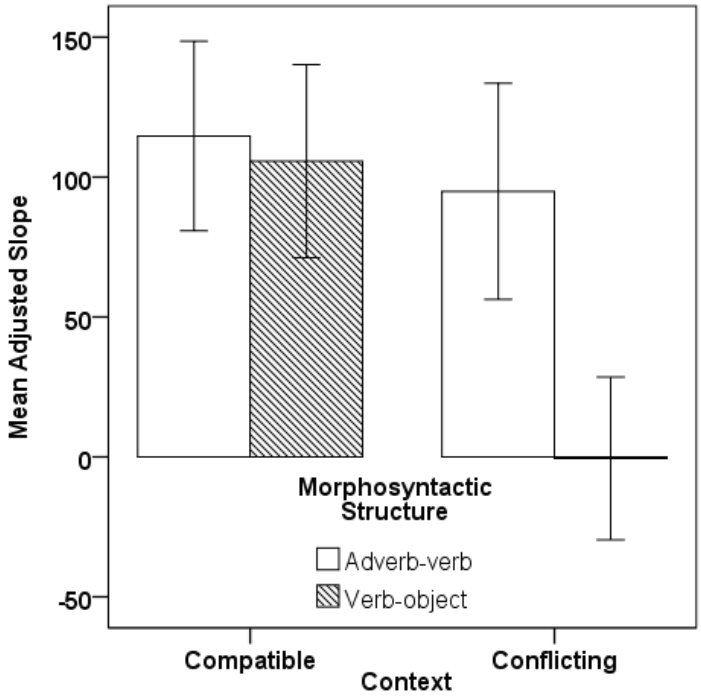


Figure 5.6: Means and error bars (±2 SE) for the adjusted slope values (right half) in leftmost position. Values broken down by context (left = compatible, right = conflicting) and structures (shaded = verb-object, white = adverb-verb).

In order to test whether the differences between the two contexts and the two prosodic structures is statistically significant, a by-subjects Repeated Measures (RM) ANOVA was conducted, with context and prosodic structures as the two main factors, and the adjusted slope values as the dependent variable.

Both the factor PROSODIC STRUCTURE [$F(1,12) = 7.36$, $p = 0.019$] and the factor CONTEXT [$F(1,12) = 11.59$, $p = 0.005$], as well as their interaction

[PROSODIC STRUCTURE*CONTEXT: $F(1,12) = 22.91$, $p < 0.001$] exert a significant influence on the tonal trajectories of the leftmost targets. Because a significant interaction was found, the two contexts were investigated separately for a difference between the prosodic structures. It turned out that the difference between the prosodic structures is highly significant in conflicting context [$F(1,12) = 34.92$, $p < 0.001$], but not significant in compatible context [$F(1,12) = 0.31$, $p = 0.59$, ns]. This confirms the impression that tones in leftmost position in verb-object structures are influenced by coarticulation to a different (greater) extent than those in adverb-verb structures.

5.3.3.2 Rightmost targets

For the tones on monosyllabic target words in rightmost position, the same slope computations were performed as for the tones in leftmost position. Figure 5.7 shows the first half of the tonal slopes in conflicting context, split between the two prosodic structures ((ADV-ADV)(**V**) = solid lines, ((V-V)(**O**)) = dashed lines) for the four lexical tones.

Figure 5.7 shows that in rightmost position, there is no systematic difference in the effect of coarticulation across all four tones. For the two rising tones (which in the first half of their trajectories have a falling slope in conflicting condition, compare Figure 5.4), the realization in adverb-verb structures seems to have reversed the tonal contour to a greater extent than the realization in verb-object structure for the low falling tone, but there is barely any difference for the high rising tones.

For the two falling tones, it seems that the tonal trajectories on the adverb-verb structures are more true to the original falling trajectories than the realization in verb-object structures. In sum, two of the tones show slightly more coarticulation for the verbs in adverb-verb structures, one of the tones shows slightly more coarticulation for the object in verb-object structure, and one of the tones shows very little difference between the two structures.

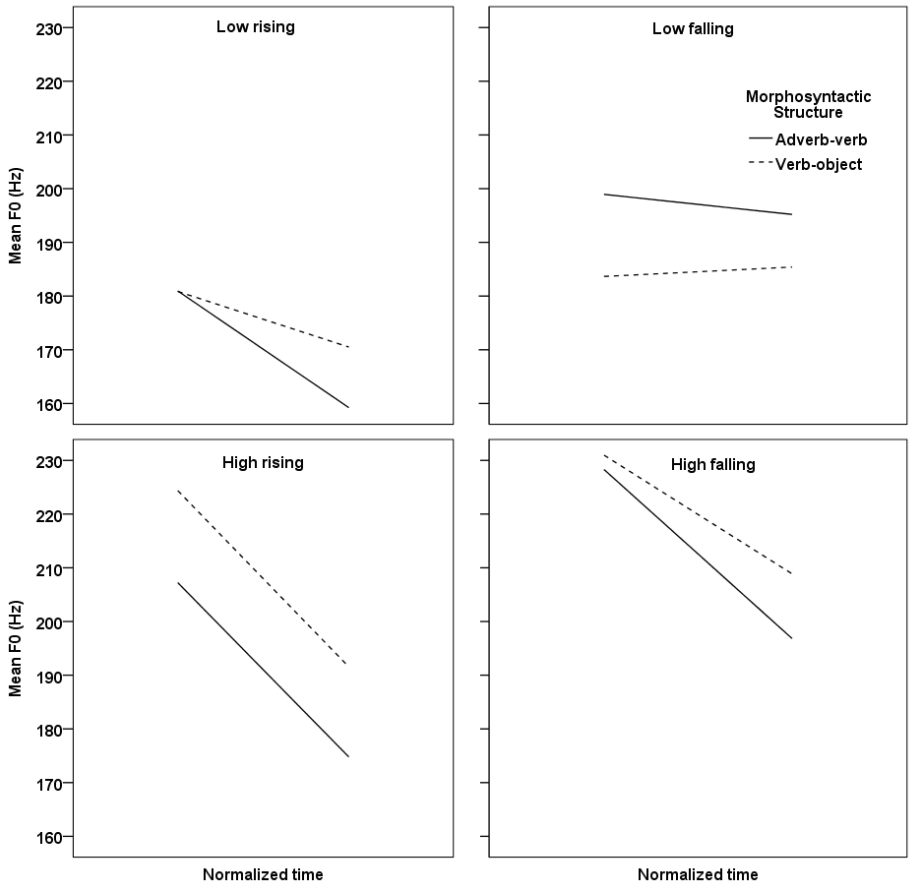


Figure 5.7: Linear representations of the tonal slopes (leftmost half) in rightmost position in conflicting context, broken down by tone on the target syllable and structure (dashed lines = verb-object, solid lines = adverb-verb).

After adjusting the slope values for the rightmost targets and pooling over all four tones in the same manner as was done for the leftmost tones, it was found that the coarticulation effects cancel each other out, because they do not systematically point in the same direction as the effects on the leftmost tones do. In other words, when pooling over all four tones, the trajectories are overall flatter in conflicting than in compatible context, but there is no big difference between the prosodic structures on either side. Figure 5.8 illustrates this.

A by-subjects RM ANOVA confirms that there is a highly significant main effect of CONTEXT [$F(1,12) = 23.98$, $p < 0.001$], but no significant effect of PROSODIC STRUCTURE [$F(1,12) = 0.02$, $p = 0.907$, ns], or an interaction between the two factors [CONTEXT*PROSODIC STRUCTURE: $F(1,12) = 1.27$, $p = 0.281$, ns]. This shows that in rightmost position, the two prosodic structures are not systematically different in slope, which speaks for a coarticulation effect that is equally strong across the two prosodic structures.

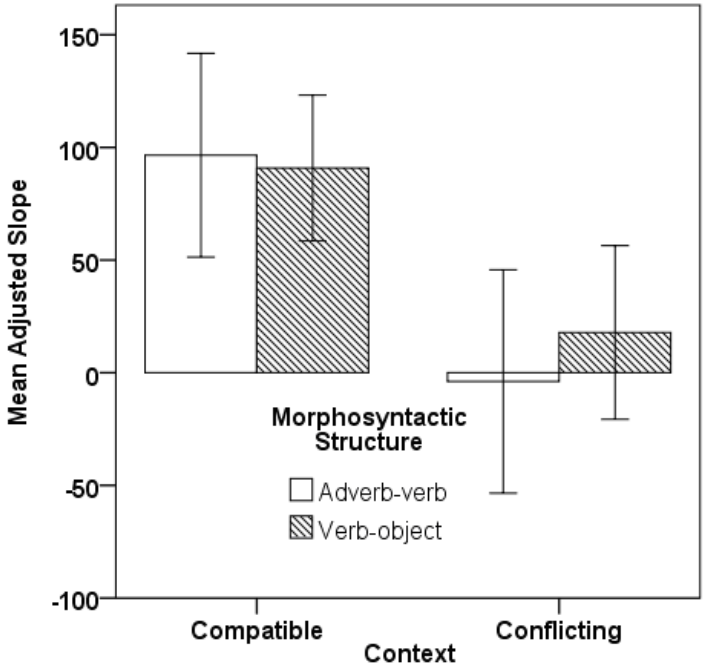


Figure 5.8: Means and error bars (±2 SE) for the adjusted slope values (left half) in rightmost position. Values broken down by context (left = compatible, right = conflicting) and structures (shaded = verb-object, white = adverb-verb).

5.3.4 Duration

In the above statistics, the factor duration is already included, in so far as the slope measurements have been calculated using the duration data for the individual conditions. However, it is also interesting to look at the duration data itself, since it can give insights into the type of tonal modification that occurs in the different contextual and prosodic conditions. Specifically, it can be

investigated whether the relative steepness/slope of the tonal contours is directly covariant with lengthening, or whether the slopes are modified independent of the durational modification. Figure 5.9 represents the duration results graphically.

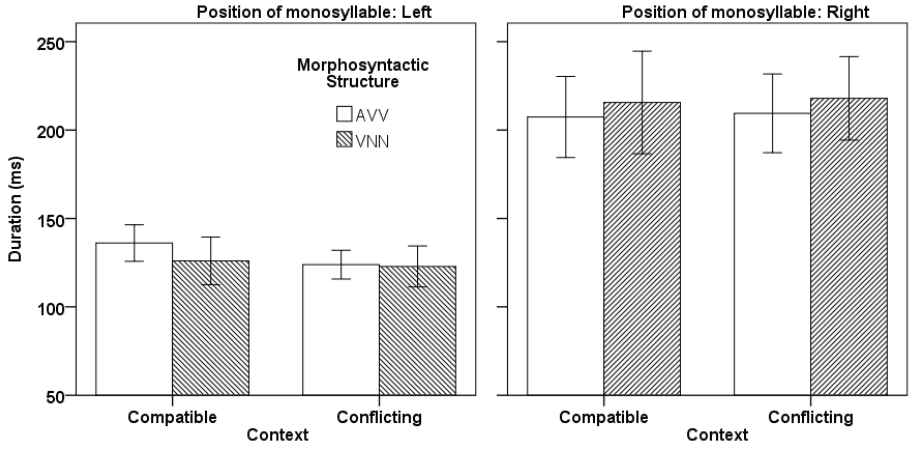


Figure 5.9: Means and error bars (±2 SE) for the duration values in leftmost (left graph) and rightmost (right graph) position. Values broken down by context (left = compatible, right = conflicting) and structures (shaded = verb-object, white = adverb-verb).

The duration of the target syllables was compared in a by-subjects RM ANOVA, with POSITION, CONTEXT, and PROSODIC STRUCTURE as main factors. The results show that of the three factors, only POSITION (monosyllabic target syllable leftmost/rightmost) had a significant effect on the duration of the target syllable [$F(1,12) = 35.17$, $p < 0.001$], whereas neither CONTEXT (conflicting/ compatible) [$F(1,12) = 1.01$, $p = 0.334$, ns] nor PROSODIC STRUCTURE (ADV-V vs. V-O) [$F(1,12) = 0.09$, $p = 0.765$, ns] exerted a significant effect.

The strong effect of POSITION meant that also the interaction POSITION*CONTEXT reached significance [$F(1,12) = 5.04$, $p = 0.044$], and the interaction POSITION*PROSODIC STRUCTURE approached significance [$F(1,12) = 3.94$, $p = 0.07$]. Neither the interaction CONTEXT*PROSODIC STRUCTURE [$F(1,12) = 3.04$, $p = 0.107$, ns] nor the three-way interaction POSITION*CONTEXT* PROSODIC STRUCTURE [$F(1,12) = 0.77$, $p = 0.399$, ns] reached significance.

Because there was a significant interaction POSITION*CONTEXT, it is worthwhile to examine the two positions separately for effects of the two other factors. For the targets in rightmost position within the phrase, neither the factor CONTEXT [$F(1,12) = 0.27$, $p = 0.613$, ns] nor the factor PROSODIC STRUCTURE [$F(1,12) = 1.23$, $p = 0.29$, ns] nor their interaction [CONTEXT*PROSODIC STRUCTURE $F(1,12) = 0.002$, $p = 0.961$, ns] reached significance. For the targets in leftmost position within the phrase, only the factor CONTEXT [$F(1,12) = 8.6$, $p = 0.013$] reached significance, while neither the factor PROSODIC STRUCTURE [$F(1, 12,) = 3.59$, $p = 0.082$, ns] nor the interaction between the two factors [CONTEXT*PROSODIC STRUCTURE $F(1,12) = 2.71$, $p = 0.126$, ns] reached significance.

5.4 Discussion: Tonal coarticulation and prosodic structure

The foregoing experiment tested whether prosodic structure influences tonal coarticulation in Wenzhou Chinese, and if so, whether prosodic boundaries or prosodic headedness play a larger role. In order to quantify the extent of the influence of neighboring tones onto the tonal trajectories of the target words, their realizations were compared between compatible context, in which no adjustment of tonal trajectories is expected, and conflicting context, in which the tonal trajectories should show influence from the adjacent tonal context.

As has been shown in the previous section, the tonal slopes were indeed significantly flatter in verb-object structures in leftmost position in conflicting context, compared with compatible contexts. In contrast, the adverb-verb structures showed similar tonal slopes in both conflicting and compatible context in leftmost position. This speaks for a greater autonomy of the targets in adverb-verb structures in their tonal realization, whereas the targets in verb-object structures were subject to significant influence from the neighboring tonal context.

Up to this point, the results are compatible with both hypothesis (ii) and hypothesis (iii) from section 5.1.4. In order to decide whether the observed effect is brought about by the prosodic boundary strength between the target word and its tonal context, or between the prosodic headedness distribution within the prosodic structure, the results in leftmost position have to be compared with the results in rightmost position. As discussed in the previous section, in rightmost position, the target monosyllables in both structures are implemented in a similar way in both compatible and in conflicting context. This leads to the interpretation that it is not prosodic boundary strength which

determines the amount of coarticulation in the target words of this experiment, since this boundary strength is identical for target words in leftmost and rightmost position.

Rather, the results lead to conclude that the prosodic headedness of the target words is responsible for the differences in magnitude of coarticulation between the two prosodic structures. As laid out in section 5.1.4, in terms of prosodic headedness, both structures display a prosodically strong position on the right side, but only the adverb-verb structures also have a prosodically strong position on the left side. Therefore, finding a difference in tonal coarticulation in leftmost but not in rightmost position in the phrases strongly suggests that the effect is dependent on prosodic headedness, rather than on prosodic boundary strength.

Taking into account the durational data, it is interesting to note that the difference in tonal slopes in the leftmost tones did not correlate with a significant duration difference between the two structures. This indicates that the tonal implementation effect is not just due to e.g. greater pre-boundary lengthening in prosodic phrase compared to prosodic word context, where the longer duration of the syllable would allow a fuller implementation of the tone, compared to the truncated tonal realization in non-lengthened positions. Rather, the duration data shows that, while there was a slight, non-significant lengthening in the adverb-verb structures compared to the verb-object structures (see Figure 5.9), this difference alone cannot be responsible for the significant slope effect.

5.5 Tonal coarticulation under narrow focus

If tonal coarticulation in trisyllabic phrases is indeed influenced by prosodic structure and in particular by prosodic headedness, an interesting follow-up question to ask is: what happens to this effect under narrow focus? As laid out in section 5.1.3, research for other Chinese dialects has suggested that focus exerts a strengthening effect on tonal realization, which leads to magnified implementation of tonal contours and to greater resistance of focused tones to influence from neighboring tonal contours.

For Wenzhou, it has been suggested that the effect of focus is therefore best accounted for as prosodic prominence effect. Under focus, the syntactically derived prosodic structure is overridden, and the focused constituent becomes the prosodic head of the entire intonation phrase (Chen 2000: 511). Similar accounts have been proposed for other languages, in order to explain the effects

of focus in an indirect way via its proposed influence on prosodic prominence (Büring 2010; Gussenhoven 1992; Truckenbrodt 1999). The idea is to relate focus to maximal prominence in its prosodic domain, and in cases where focus requirements collide with the syntactically derived prosodic structure, the prosodic structure is changed to fulfill the focus requirements.

For the results presented in the foregoing section, such an interpretation of focus would lead to the assumption that under focus, the effects of prosodic structure should be neutralized. More specifically, if the focused constituent is assumed to be the prosodic head, regardless of which constituent should be the prosodic head according to the morphosyntactic structure, it can be expected that any effect that is brought about by the morphosyntactic structure should disappear in the presence of focus.

5.5.1 Stimuli, speakers, experimental procedure, data analysis

In order to test this hypothesis, a second experiment was conducted, using the same materials as the first experiment. In the second experiment, however, the trisyllabic phrases were presented in the context of an alternative question, which induced narrow contrastive focus on the monosyllabic target word, as exemplified in (4).

(4) Q: 到 中国 或者 走 中国?
tɜ42 tɕoŋ35.kai42 va22.tsɜ33 tsau35 tɕoŋ35.kai42
reach China or walk-to China
''To reach China' or 'to walk towards China'?'
A: 到 中国
tɜ42 tɕoŋ35.kai42
reach China
'To reach China.'

Similarly to the first experiment, the stimulus phrase in the answer constitutes its own utterance, and therefore its realization should, for all intents and purposes, be comparable to the realization of the stimuli in the first experiment, apart from the additional influence of contrastive focus on the target monosyllable.[20] The

[20] In the following, the results of the second experiment reported in section 5.5, which deals with narrow contrastive focus, will be compared to the results of the first experiment discussed in sections 5.2-5.4, which presented the items without a specific information-structural context. It is a matter of debate whether a presentation out of the

same 19 speakers as in the first experiment took part in the second experiment, but in order to further increase the comparability of the two experiments, only the data from the same 13 speakers as in the first experiment was analyzed and will be reported. Each of the speakers recorded the trisyllabic phrases under focus after completing the recording of the trisyllabic phrases in isolation, so the speakers were already familiar with the materials.

The recording procedure for the trisyllabic phrases under focus was identical to the procedure used to obtain the recordings in isolation (see section 5.2.3). After excluding incomplete and erroneous renditions, 1265 target tokens remained for analysis. These were segmented and labeled in the same way as the target tokens from the first experiment, and subjected to the same F_0 and duration data extraction procedure (see section 5.2.4). Ninety-three tokens with utterance-final breathiness/creakiness were excluded from the data measurements, and the automatic F_0 measurements were hand-corrected for tracking errors such as octave jumps in eight cases.

5.5.2 Results

In order to facilitate the comparison to the results of the first experiment, the graphs representing the realizations of the monosyllabic target words under focus will be set up in the exact same way as the graphs in section 5.3. All data was transformed in the same way (for example, for the following to graphs, the measurements have been scaled to the individual speaker's pitch ranges before being averaged across speakers).

5.5.2.1 Position

Figure 5.10 represents the tonal trajectories of the monosyllabic words under contrastive focus, averaged across repetitions, speakers, contexts, and prosodic structures, and split by tones and position within the trisyllabic phrase. As can be seen, the general trend from the unfocused data can also be confirmed for the focused data, namely that the tonal trajectories are more pronounced and cover a

---

blue evokes so-called "broad" or neutral focus that is comparable to an "unfocused" condition, or whether it just puts the entire utterance in focus, as argued for example in Lambrecht 1994. Since the current experiment is concerned with narrow contrastive focus on one of the three syllables in the trisyllabic phrases, the isolated context will be referred to as "non-focused" or "unfocused" for ease of comparison. However, this should not be taken to imply the complete absence of any (broad) focus in the isolated examples in the first experiment.

wider area of the speakers' pitch ranges in rightmost compared to leftmost position.

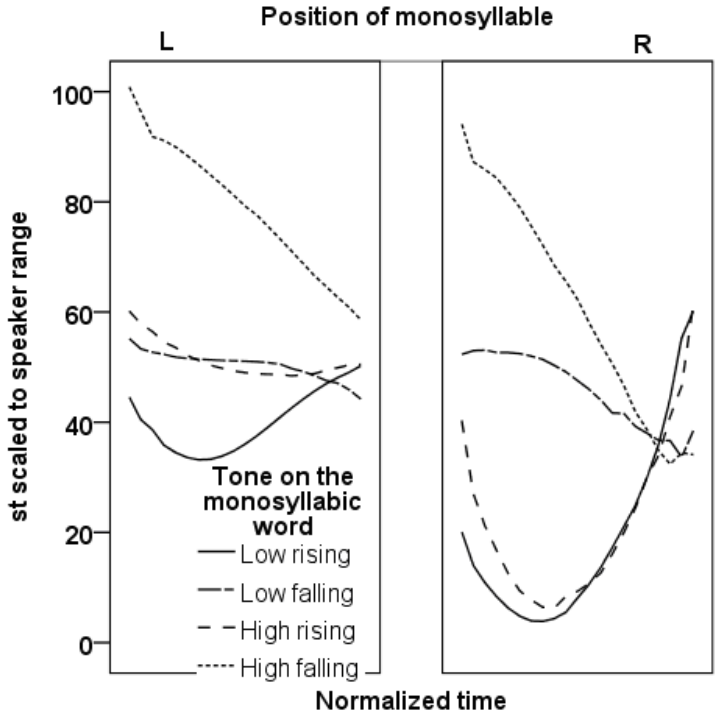


Figure 5.10: Averaged and time-normalized F_0 contours of the four lexical tones under narrow focus, broken down by position (L = leftmost, R = rightmost).

5.5.2.2 Context

Because of the above-mentioned split between tonal realizations in leftmost vs. rightmost position, the following analysis will again present the results for the both positions separately. The graph on the left side of Figure 5.11 shows the tonal trajectories of the tones in leftmost position, split by conflicting (top) and compatible (bottom) context, and the graph on the right shows the same for the tonal trajectories in rightmost position.

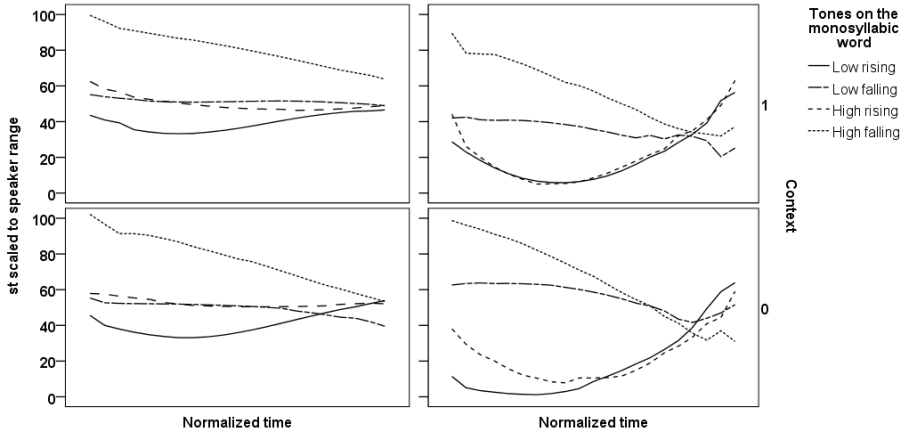


Figure 5.11: Averaged and time-normalized F_0 contours of the four investigated tones in initial (left) and final (right) position, broken down across contexts.

Again, the differences between the tonal realizations in the two contexts in focused condition mirror the effects found in the unfocused condition. The difference between the trajectories in the two contexts again pertains mainly the half of the contours that is immediately adjacent to the context tones, namely the right half for the leftmost tones and the left half for the rightmost tones. The effects for the specific tones are also comparable: rising tones are largely deprived of their final rising portion in leftmost position, and start higher in rightmost position. Falling tones are also flattened in leftmost position in conflicting context, and start from a less high starting point in rightmost position in conflicting context compared to compatible context.

5.5.2.3 Prosodic structure

The same calculations were performed on the data as in the first experiment to visualize the influence of the two prosodic structures onto tonal coarticulation. Specifically, the slope values were computed for the right half of the leftmost tones, and for the left half of the rightmost tones, to zoom into the area that is most affected by tonal coarticulation.

5.5.2.3.1 Leftmost targets

As Figure 5.12 shows, the difference for the leftmost trajectories between the two prosodic structures that could be seen in the unfocused data cannot be found

back in the data under focus. Apart from a slight difference in the low rising structure, the two slopes run more or less parallel. This indicates that there is no significant difference in slope between the two prosodic structures for leftmost targets.

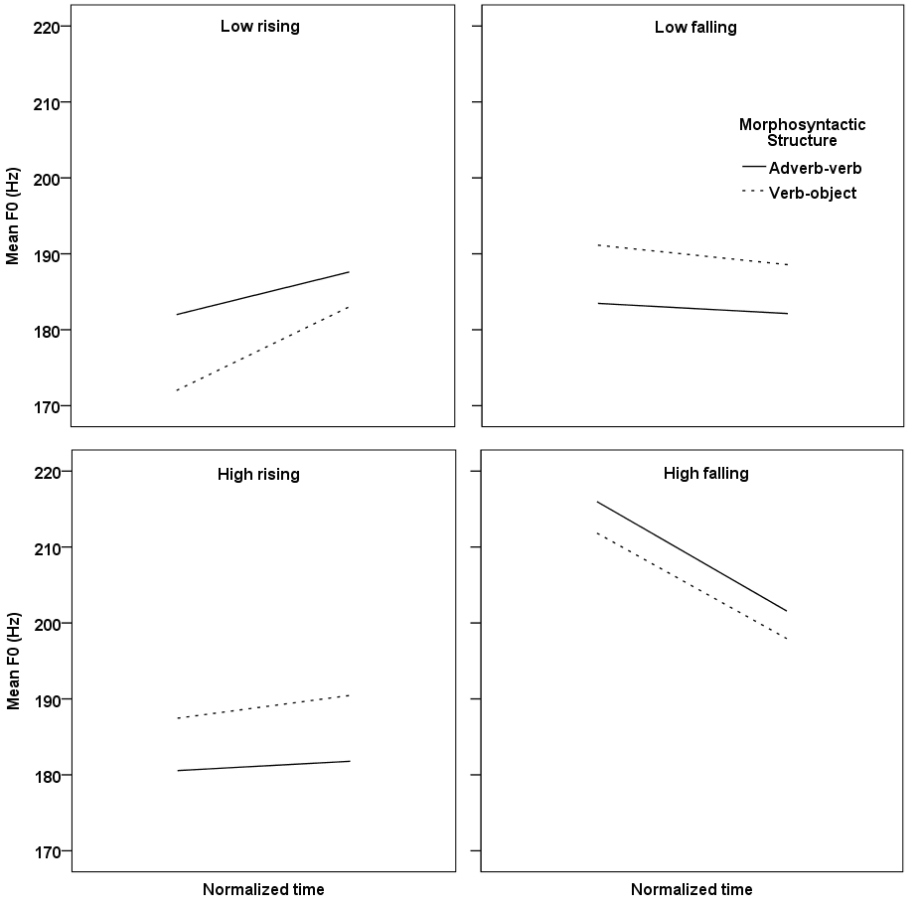


Figure 5.12: Linear representations of the tonal slopes (rightmost half) in leftmost position in conflicting context under focus, broken down by tone on the target syllable and structure (dashed lines = verb-object, solid lines = adverb-verb).

For statistical comparison, the slope values were adjusted and averaged across the four tones, as shown in Figure 5.13. A by-subjects RM ANOVA on the adjusted slope values shows that the only significant difference is brought about by the factor CONTEXT [$F(1,12) = 14.87$, $p = 0.002$], while both PROSODIC STRUCTURE [$F(1,12) = 3.6$, $p = 0.82$, ns] and the intercept between the two factors [CONTEXT*PROSODIC STRUCTURE: $F(1,12) = 0.92$, $p = 0.356$, ns] turn out to not be significantly different from each other. This confirms what the inspection of the graphical slopes already led to assume: the effect of prosodic structure upon the magnitude of coarticulatory adjustment in the tonal realization on initial target words disappears under contrastive focus.

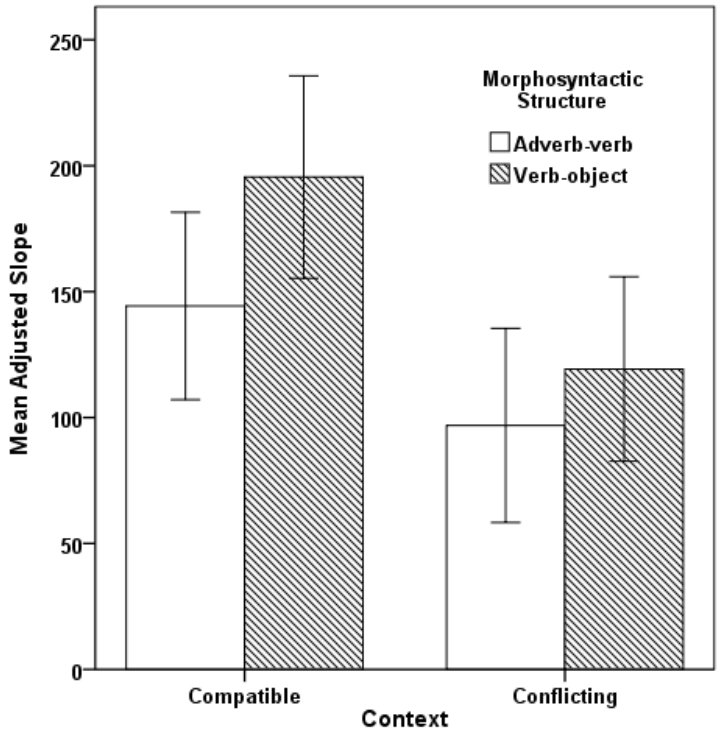


Figure 5.13: Means and error bars (±2 SE) for the adjusted slope values in leftmost position under contrastive focus. Values per tone broken down by context (left = compatible, right = conflicting) and structures (shaded = verb-object, white = adverb-verb).

At the same time, the Figure 5.13 shows an important difference between the leftmost targets in the first and in the second experiment. In the first experiment, the slopes in conflicting contexts were almost flat for the verb-object structures. In contrast, in the second experiment, the tonal trajectories are relatively steep in both conflicting and compatible context. There is still a difference between the two contexts, and as in experiment 1, it is a significant one, in that the tones in compatible context are steeper across the board than the tones in conflicting context. However, under focus, the targets are strengthened in their tonal realization in both contexts, in that the slopes even in compatible context are relatively steeper under focus than in the first experiment (compare Figures 5.6 and 5.13).

5.5.2.3.2 Rightmost targets

Figure 5.14 illustrates the tonal slopes in rightmost position under focus. As in the non-focused condition, the tones on the right side show no consistent influence of prosodic structure onto tonal realization. For the two rising tones, the tones in verb-object structure seem to be less falling and consequently less influenced by the preceding context than the tones in adverb-verb structure. For the falling tones, the picture is reversed. Here, it is the adverb-verb structures that show steeper falling contours and thereby less influence from coarticulation. In sum, the picture that was shown in non-focused condition is repeated, namely no systematic difference in slope between the two prosodic structures.

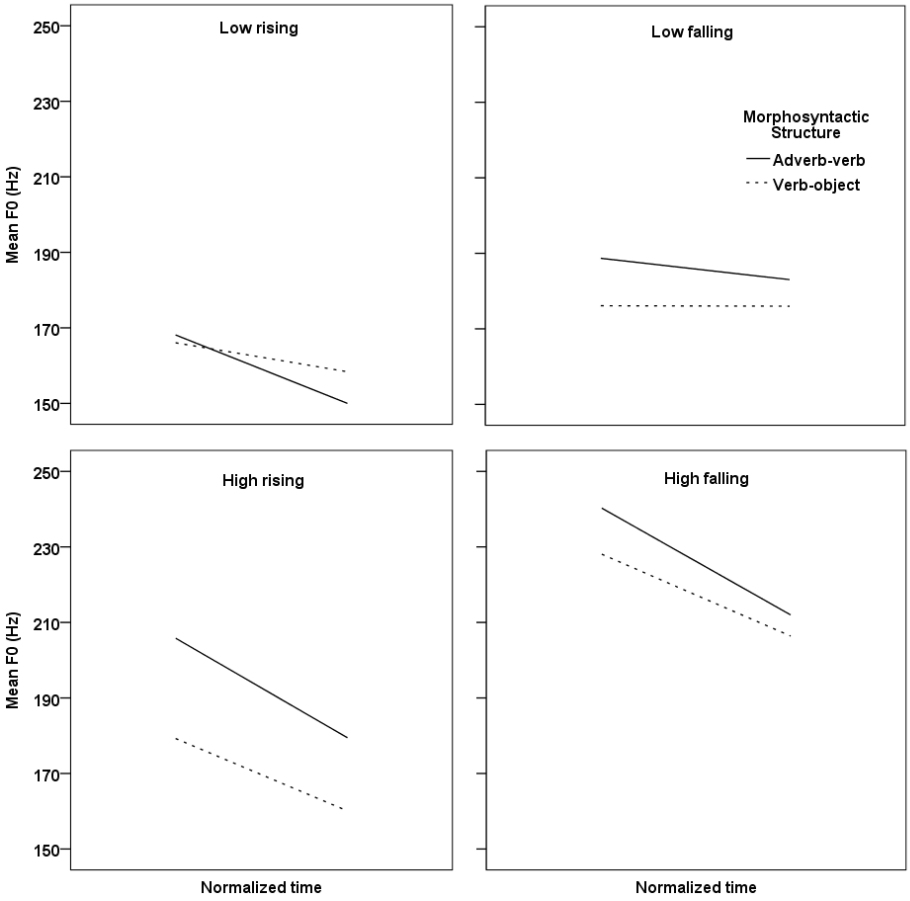


Figure 5.14: Linear representations of the tonal slopes (leftmost half) in rightmost position in conflicting context under focus, broken down by tone on the target syllable and structure (dashed lines = verb-object, solid lines = adverb-verb).

Also for the tones in rightmost position, the slope values were adjusted and then averaged across tones, as shown in Figure 5.15. A by-subjects RM ANOVA shows that neither CONTEXT [$F(1,12) = 3.19$, $p = 0.1$, ns] nor PROSODIC STRUCTURE [$F(1,12) = 1.17$, $p = 0.301$, ns] or the interaction between the two factors [CONTEXT*PROSODIC STRUCTURE: $F(1,12) = 0.03$, $p = 0.866$, ns] induces a significant difference in the slopes of the tonal realizations.

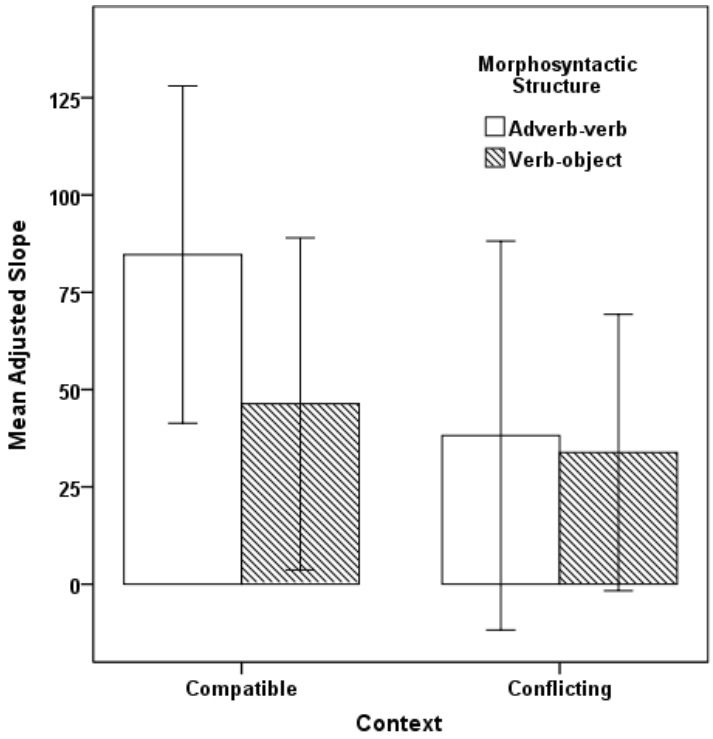


Figure 5.15: Means and error bars (±2 SE) for the adjusted slope values in rightmost position under contrastive focus. Values per tone broken down by context (left = compatible, right = conflicting) and structures (shaded = verb-object, white = adverb-verb).

5.5.2.4 Duration

As in the previous experiment, the duration measurements of the target monosyllables under focus were averaged over speakers, and then tested in a by-subjects RM ANOVA. Figure 5.16 illustrates the duration results under focus graphically.

Similar to the previous experiment, only the factor POSITION (leftmost/rightmost within the phrase) exerts a significant influence on the duration data [$F(1,12) = 135.95$, $p < 0.001$], whereas neither the factor CONTEXT (conflicting/compatible) [$F(1,12) = 0.11$, $p = 0.744$, ns] nor the factor PROSODIC STRUCTURE (ADV-V vs. V-O) [$F(1,12) = 1.65$, $p = 0.223$, ns] exerted a

significant difference on their own. Furthermore, a significant two-way interaction CONTEXT*PROSODIC STRUCTURE [$F(1,12) = 17.78$, $p = 0.001$] was found, as well as a significant three-way interaction POSITION*CONTEXT* PROSODIC STRUCTURE [$F(1,12) = 8.93$, $p = 0.011$]. None of the other interactions reach significance [POSITION*PROSODIC STRUCTURE ($F(1,12) = 0.38$, $p = 0.55$, ns), POSITION*CONTEXT ($F(1,12) = 0.34$, $p = 0.569$, ns].

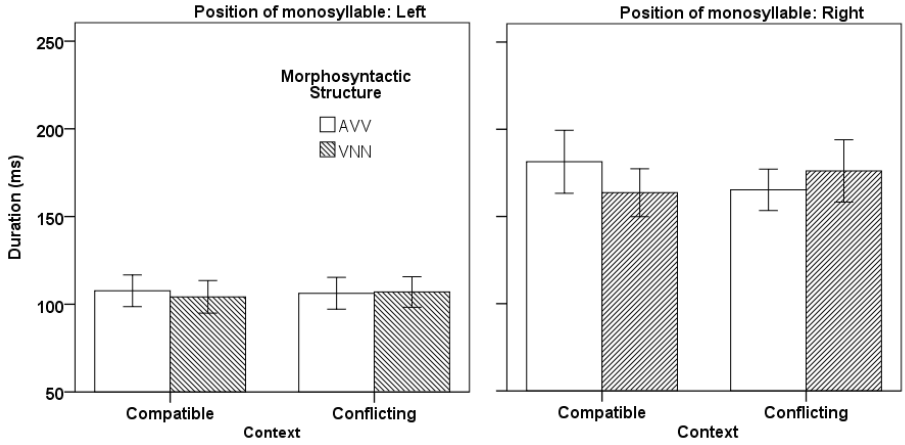


Figure 5.16: Means and error bars (±2 SE) for the duration values in leftmost (left graph) and rightmost (right graph) position under focus. Values broken down by context (left = compatible, right = conflicting) and structures (shaded = verb-object, white = adverb-verb).

A second RM ANOVA compared the duration results from the first experiment (recording in isolation) directly with those of the second experiment (recording under focus), with the added factor FOCUS as a within-speaker variable. First of all, the factor FOCUS was found to induce a significant main effect between the duration measurements of the first experiment and those of the second experiment [$F(1,12) = 55.33$, $p < 0.001$]. Closer inspection of the duration values of the target syllables under focus and in isolation showed however that the values were actually lower in focus condition than in isolation.

This can be explained by the presence of the focus-inducing contrastive sentence in the second experiment, which preceded the target phrases and led to a higher on-average speech rate across speakers. Comparison of the average duration of the entire trisyllabic phrases confirms this assumption: in isolation,

the average duration of the trisyllabic phrases is 0.74 seconds (SD: 0.11 seconds), whereas the average duration of the same trisyllabic phrases in the focus condition is 0.6 seconds (SD: 0.09 seconds). At the same time, the ratio between the duration of the target monosyllable in relation to the duration of the entire trisyllabic phrase is similar in both experiments, namely 23% (isolation: 0.17 seconds, SD: 0.05 seconds, focus: 0.14 seconds, SD: 0.04 seconds). Therefore, the significant effect of FOCUS on the duration data can actually be explained by the speech rate, which masks any focus-induced lengthening that might be present on the target syllables in the second experiment.

Furthermore, the factor POSITION (leftmost/rightmost within the phrase) also exerted a significant main effect on the duration data when comparing the first with the second experiment [$F(1,12) = 66.17$, $p < 0.001$]. However, the two-way interaction FOCUS*POSITION turned out to be not significant [$F(1,12) = 2.81$, $p = 0.119$, ns], which shows that the effect of position is the same in the two experiments (namely longer duration of the target syllables in rightmost than in leftmost position). The factor PROSODIC STRUCTURE did not exert a significant difference on the duration measurements when comparing the first to the second experiment [$F(1,12) = 0.006$, $p = 0.94$, ns], nor did the factor CONTEXT exert a significant difference on its own [$F(1,12) = 1.56$, $p = 0.186$, ns].

In terms of interaction, only two factor combinations reached significance. A significant effect of the two-way interaction PROSODIC STRUCTURE*CONTEXT [$F(1,12) = 15.08$, $p = 0.002$] indicates that there is a difference in the duration measurements between the first and the second experiment that depends on prosodic structure in covariance with context. The three-way interaction FOCUS*POSITION*PROSODIC STRUCTURE [$F(1,12) = 5.31$, $p = 0.04$] indicates that the duration measurements differ between the two experiments when taking position and prosodic structure into account in combination. No other interaction reached significance.

5.5.3 Discussion: Tonal coarticulation under focus

The results of the first experiment showed that tonal implementation and its susceptibility to influence from neighboring tones in Wenzhou were influenced by prosodic structure, specifically by prosodic prominence in prosodic head positions. It was shown that tones in leftmost position in the target phrases, in which there is a difference in prosodic headedness between the two prosodic structures, showed a different amount of influence from neighboring tonal targets, with the tones in prosodically weak positions coarticulating more with the tonal context than tones in prosodically strong positions. In contrast, the

tones in rightmost position in the target phrases showed no such difference between the two prosodic structures, which was explained with the fact that the rightmost syllables are prosodically strong in both contexts.

The second experiment tested whether this effect of prosodic strength could be influenced by focus, and specifically, whether the effect of prosodic structure would simply be nullified under focus, or whether focus would induce its own strengthening, even on those targets that were already in prosodically strong positions. The results presented in the foregoing sections point in the direction of the latter assumption.

The second experiment found that under focus, the difference in tonal slopes between the two prosodic structures that was found in the first experiment disappeared. Both the targets in rightmost and leftmost position showed similar slopes for the adverb-verb structures and for the verb-object structures. A superficial examination of these findings could lead to the assumption that under focus, the prosodic structure is changed to reflect the focus structure, and that focus takes over the task of assigning prosodic prominence and thereby overrides the prosodic structure that is built on basis of the morphosyntactic structure.

However, closer inspection of the results of the second experiment, and a comparison with those of the first experiment, shows that such an account cannot explain all the findings of the current data. For the targets in leftmost position, it can be seen that the slope values in compatible context are higher for both prosodic structures under focus compared to isolation. If focus were simply a re-allocation of prosodic headedness, it would not be expected that it should make a difference in the implementation of targets that are not in conflict with their tonal environment.

Even more so, a comparison between the tonal slopes in rightmost position in the two experiments also shows a clear effect of focus. In both experiments, there is no difference in prosodic headedness between the two prosodic structures in rightmost position, since both structures have a prosodic prominence on the right side. Nonetheless, there is a significant difference between tones in conflicting and compatible context even in rightmost position, with the tones in both structures being influenced by the preceding tonal context in conflicting position.

Under focus, however, context ceases to affect the implementation of tones in these prosodically strong positions. As Figure 5.15 shows, focus further strengthens the implementation of tones, even if they already are in prosodically prominent positions, so that they are realized with relatively greater autonomy

from the conflicting tonal context. Under an account which simply interprets focus as prosodic prominence allocation, this finding would be completely unexpected. If focus simply shifts prosodic headedness, tones that already are in a prosodic head position morphosyntactically would be predicted to be unaffected by focus. In other words, focus would be predicted to strengthen prosodically weak positions, but it would not be expected to further strengthen prosodically strong positions.

This, however, is exactly what the comparison of the findings of the two experiments indicate. Under focus, tonal implementation is strengthened across the board, even for tones in already prosodically strong positions. This means that an explanation of focus as prosodic headedness is inadequate for the findings of the two experiments presented here. Rather, the influence of prosody onto tonal coarticulation and the influence of focus onto tonal coarticulation have to be kept apart, even if the two factors induce similar effects onto the implementation of tonal contours.

5.6 Conclusion

In the present chapter, two research questions were investigated, as laid out in section 5.1.4: (i) which contextual factors affect tonal coarticulation in Wenzhou Chinese, and (ii) how do these factors interact with the strengthening of tonal implementation induced by focus? In order to test the first question, the implementation of rising and falling tones in rightmost and leftmost position in trisyllabic phrases in Wenzhou Chinese was investigated. By comparing the implementation of tonal contours in conflicting and compatible contexts, it was measured to what extent the tonal trajectories were affected by the adjacent tonal context.

In order to specify the exact nature of the contextual influence, two different prosodic structures were tested, which differed along two prosodic criteria: prosodic boundary strength and prosodic head position. Since these two criteria were non-overlapping, specific predictions could be made that would allow to test which of the two criteria was responsible for the prosody effect on tonal coarticulation.

The results of the first experiment indicated that it was prosodic headedness which could best explain the differences in tonal coarticulation between the two prosodic structures. In prosodically strong positions, it was found that tones were coarticulated relatively less with adjacent tones than in prosodically weak positions. In these weak positions, it was found that the tonal

slopes, particularly the slopes that were immediately adjacent to the tonal context, were implemented with flattened or directionally flipped tonal trajectories.

At the same time, the durational results showed that this flattening of tonal trajectories in prosodically weak position was not a direct by-product of durational truncation. Compared to prosodically strong positions, speakers took a comparable amount of time for the realization of the tonal trajectories, but implemented them in a more distinct way in strong positions.

Under focus, what could be observed is best described as a general strengthening effect that boosted the tonal implementation both in prosodically weak and in prosodically strong positions. In the weak position, this meant that the difference between the two prosodic contexts in terms of tonal coarticulation disappeared, and both prosodic structures showed a similar amount of tonal coarticulation. However, tonal implementation was also strengthened in prosodically strong positions. Across the board, under focus, the amount of difference in tonal implementation between conflicting and compatible contexts was reduced. This means that tones were realized more autonomously and independently from the adjacent tonal context, even if this context conflicted with the tonal targets, under focus.

These findings have important consequences for theoretical accounts of the interaction between prosodic structure and focus. Particularly, the current results show that focus should not be conceptualized as being implemented as prosodic prominence. The effects of prosodic prominence on tonal coarticulation, as presented in the current chapter, are independent from the effects of focus on tonal coarticulation, even if the two factors influence tonal implementation in a similar fashion.

In that respect, the current findings present an argument against an extreme version of the view that focus influences tonal realization only indirectly, via modifying the prosodic prominence status of the focused constituent (see similar proposals in Chen 2009; Féry 2010; Féry & Ishihara 2010). Rather, the current findings suggest that both focus and prosodic structure affect the strength and autonomy of tonal implementation, but do so in a separate way.

Chapter 6

Sentential F_0 scaling in Wenzhou Chinese

6.1 Introduction

6.1.1 Views on pre-planning of sentential F_0 scaling

It has been a long-standing research question in linguistics how much material speakers can (and do) pre-plan before they start uttering a sentence. One way to evaluate the amount of pre-planning from the speaker's perspective is to compare the height of initial F_0 peaks in sentences of different lengths. The basic assumption is that if speakers start higher in longer than in shorter sentences, this means that they take into account the greater number of upcoming syllables and pitch peaks, and adjust their implementation of the intonational melody of the sentence from the very beginning. This idea of an inherent link between sentence length and initial peak height as evidence of pre-planning has been put forward for languages such as Swedish (Bruce 1977; Gårding 1979), Danish (Thorsen 1980), and English (Sorensen & Cooper 1980), and is sometimes referred to as the "global hypothesis" (Prieto et al. 2006).

However, the possibility of an inherent link between sentence length and the scaling of initial F_0 peaks in a sentence has also been challenged on experimental grounds. In particular, Pierrehumbert 1979, 1980 and Liberman & Pierrehumbert 1984 argue that the lowering of successive pitch peaks in English can be modeled more accurately by assuming a constant F_0 decay from one accent to another, and that speakers may either reach their F_0 baseline before the end of the sentence or adjust the slope of the F_0 decline according to sentence length. Such a view of been referred to as the "local hypothesis" (Prieto et al. 2006).

Over the last decades, several studies in different languages have attempted to tease apart the two hypotheses (e.g. Avesani 1987 for Italian, Ladd & Johnson 1987 for English, Kubozono 1993 for Japanese, Prieto et al. 1996 for Mexican Spanish, Rialland 2001 for Dagara, Arvaniti 2003 for Greek, Laniran & Clements 2003 for Yoruba, Connell 2004 for Mambila, and Prieto et al. 2006 for several Romance languages). Most studies, however, have a comparatively small speaker pool (between two and five speakers per language). Results of these studies have also been complicated by between-speaker variation, when for

example only a subset of the speakers seemed to use a particular tactic (such as global pre-planning).

These between-speaker variations have sometimes been taken as evidence for the need of a less restrictive theory. For example, Rialland (2001) re-interpreted the mixed results of Kubozono (1993) to indicate that speakers use a mix of local and global strategies for the pre-planning of F_0 scaling in Japanese sentences. Similarly, Prieto and colleagues (2006) interpreted their mixed results for the Portuguese speakers as "soft pre-planning", and concluded that F_0 preplanning in this language is better viewed as speaker-dependent optional mechanism. What is worth noting is that in most studies on the connection between sentential pre-planning and F_0 scaling, the focus has been on the length of the whole utterance. It is however possible that sentential constituents may also exert an independent effect on the F_0 scaling. The current study sets out to address this issue further by examining how the length of the whole sentence, as well as the length of the subconstituents (such as subject and object), might affect F_0 scaling.

Apart from sentence length, an important factor that has been reported to affect sentential F_0 scaling is the syntactic complexity of the sentence. In English for example, Ladd & Johnson 1987) show that the relative depth of syntactically embedded constituents is correlated with the strength of the prosodic boundary between these constituents, which is reflected in the relative magnitude of downstep between these constituents. Supporting experimental evidence comes from Dutch (van den Berg et al. 1992), Yoruba (Laniran & Clements 2003), and German (Truckenbrodt 2002, 2007; Truckenbrodt & Féry 2003).

What remains unclear is whether sentential F_0 scaling can also be affected by the embedding of subordinate clauses within a sentence in a similar manner. Specifically, sentences which might have comparable length and linear orders of constituents on the surface may nevertheless vary in the complexity of the syntactic embedding structure (for example, a VP vs. a CP being embedded in a sentence, both resulting in the surface word order SVVO). Our second goal is therefore to examine how the complexity of syntactically embedded structures may affect F_0 scaling.

6.1.2 Scaling of F_0 peaks and valleys in Chinese

There have been a few studies related to the issue of sentential F_0 scaling in Chinese, all of which were concerned with Mandarin. For example, Yuan 2004) and Yang & Wang 2002) investigated the declination of the F_0 minima ("baseline") to illuminate within-sentence lowering. Shih (2000) focused on the

relationship between F_0 peaks and sentence length, and reports a ternary split in the realization of the initial F_0 target in short, medium, and long sentences in Mandarin, but concludes that the intermediate differences are too small to be statistically relevant.

However, several other possible factors which might influence sentential F_0 scaling were not accounted for in Shih's study. Apart from the "narrow focus" she induced by prefacing the target sentences with wh-questions, she also varied the names of the subject referents between conditions. This makes it likely that her subjects might have interpreted the subject referents as contrastively focused even in the condition she labeled as "unmarked reading style" (Shih 2000: 247). Therefore, it is difficult to determine from her experiments the exact influence of constituent length on F_0 scaling in Mandarin, without taking the focus confound into account.

More recently, Wang & Xu (2011) took a step further and investigated the effects of several factors on F_0 scaling in Mandarin, while controlling the possible influence of focus more carefully. The sentences in Wang & Xu 2011 are varied in length by adding one or more modifiers before the object noun, but all sentences have the basic structure SVO. Similar to Shih (2000), they report no significant difference between the height of the initial F_0 peak in short, medium, and long sentences. Two problems remain in the design of Wang & Xu 2011. First, all sentences were elicited under one of the different topic and focus conditions. Therefore, it is unclear how sentence length alone might have affected the F_0 scaling in individual words. Second, Wang & Xu (2011) varied the length of the object across conditions, but measured only the F_0 peak of the subject to infer the effect of declination. Therefore, they might have overlooked a more local effect of the different object lengths, e.g. on the first peak of the object. In other words, they confounded different effects of sentence length and constituent length on F_0 scaling.

The current study will investigate another Chinese dialect - Wenzhou Chinese, which exhibits interesting prosodic differences from Mandarin. Wenzhou Chinese is a southern Wu dialect spoken in the city of Wenzhou in Zhejiang Province. It is known for its complex phonology of tone sandhi, the implementation of which is dependent on the intricate interplay between syntactic coherence, prosodic prominence, and length of the constituents (Chen 2000). This chapter will investigate how, on top of tone sandhi, tonal realization (such as F_0 scaling) is conditioned in the language and to what extent, Wenzhou Chinese might differ from Mandarin Chinese in this respect.

6.1.3 Research questions

To test the influence of sentence length and constituent length on F_0 scaling separately, the stimulus sentences in the current experiment varied the number of words in the subject and the object constituent independently. To assess the connection between sentential F_0 scaling and the complexity of syntactic structures, three different types of embedded complements (VP, IP, and CP complements) with comparable sentence and constituent length were recorded. Worth noting is that the focus structure of the test sentences was carefully controlled, so as to exclude the possible confounding effect of focus on F_0 scaling and sentence planning.

The specific research questions are recapitulated in the following:

(i) How does the length of the constituents in a sentence affect F_0 scaling in Wenzhou Chinese? Particularly, how does the length of the constituents determine the scaling of the initial peaks within these constituents, and (how) does it interact with the scaling of the F_0 valleys?

In order to answer the first research question, the length of the stimulus sentences was varied systematically by adding words to the subject and object constituent independently. Looking at the height of the first subject peak will allow us to determine whether only one of the two independent factors (subject length or object length) has an effect on F_0 peak scaling, or whether both factors interact. As for declination of F_0 minima, it will be assessed whether sentence length interacts with F_0 scaling at all, and if so, whether the length of the constituents exerts an independent effect.

(ii) How does the complexity of complex sentences affect the F_0 scaling in Wenzhou Chinese? Particularly, in sentences with similar surface order and length that differ in underlying syntactic complexity, is the difference in complexity reflected in the F_0 scaling?

To address this latter question, two different types of comparison will be made. In a first step, sentences with the surface structure SV(VO) will be investigated, which may contain either an embedded VP complement ($SV(VO)_{VP}$) or an embedded IP complement with unrealized (optional) subject ($SV(proVO)_{IP}$). In a second step, sentences with the surface structure SV(SVO) will be investigated, which may contain either an embedded IP with optional realized subject ($SV((pro/S)VO)_{IP}$), an embedded IP with obligatory subject ($SV(SVO)_{IP}$), or an

embedded CP ($SV(SVO)_{CP}$). In both investigations, the goal is to determine whether the syntactic difference is reflected in the F_0 scaling of the tonal contours, e.g. in a steeper prosodic embedding of the more complex structures compared to the less complex structures.

6.2 Methods

6.2.1 Stimuli

The stimulus material was composed in a way that it displayed consistent tonal patterns throughout the sentences. In Wenzhou Chinese, a regular tone change process (tone sandhi) affects both syllables in disyllabic compound words, and changes the tone trajectories to specific tone sandhi contours (Chen 2000). All stimuli sentences were therefore exclusively composed of disyllabic words with the same rise-fall tone sandhi contour.[21] This contour results from a combination of any tone on the first syllable with a dipping tone on the second syllable. The investigation of a tonal contour that covers several combinations of lexical tones allowed for a fairly large database of possible words, and consequently for the composition of semantically acceptable stimulus material.

To investigate F_0 scaling in SVO structures, three lexically different sentences were used for recording, and every sentence was varied between three and seven words in length by adding words to the subject and/or object constituent. This results in 27 target sentences per speaker. The three lexical sentences are given in (1), with parentheses indicating the constituent structure and different lengths of constituents.

[21] This is also the reason why a commonly applied test for downstep vs. declination is not available here, namely to compare the downstepping of HL-tone sequences with that of H-tone sequences. Wenzhou does not have a disyllabic tone sandhi contour consisting of H-tones only, and creating sentences consisting only of monosyllabic words would be highly unnatural.

(1)

a.	(美国	(大学	(同学)))	学习
	mɛi kai	*da ɦu*	*doŋ ɦu*	*ɦu zai*
	American	university	classmate	learn
	(重要	(算术	(公式)))	
	dʑuɔ ʔi	*sɵ jø*	*koŋ sɛi*	
	important	arithmetic	formula	

'The (American (university (classmate))) learns the (important (arithmetic (formula))).'

b.	(美国	(大学	(同学)))	登录
	mɛi kai	*da ɦu*	*doŋ ɦu*	*taŋ lu*
	American	university	classmate	enroll
	(重要	(大学	(目录))).	
	dʑuɔ ʔi	*da ɦu*	*mu lu*	
	important	university	register	

'The (American (university (classmate))) enrolls in the (important (university (register))).'

c.	(美国	(大学	(同学)))	翻录
	mɛi kai	*da ɦu*	*doŋ ɦu*	*fa lu*
	American	university	classmate	copy
	(新式	(外国	(音乐))).	
	saŋ sɛi	*va kai*	*jaŋ lu*	
	modern	foreign	music	

'The (American (university (classmate))) copies the (modern (foreign (music))).'

Additionally, to investigate the effect of syntactic embedding on F_0 scaling, all three sentences mentioned above were also recorded in five different types of complex sentences. For sentences with the surface structure SV(VO) and a sentence length of four words, two different embedding strategies were compared:

- $SV(VO)_{VP}$ with VP complement (embedding verb "to plan")
- $SV(\text{PRO}VO)_{IP}$ with optional unrealized embedded subject (embedding verb "to promise")

For sentences with the surface structure SV(SVO) and a sentence length of five words, three different embedding strategies were compared:

- SV((PRO/S)VO)$_{IP}$ with optional realized embedded subject (embedding verb "to promise")
- SV(SVO)$_{IP}$ with obligatory embedded subject (embedding verb "to encourage")[22]
- SV(SVO)$_{CP}$ with embedded clause (embedding verb "to remember")

Examples with the sentence from (1a) can be found in (2).

(2) a. 同学 计划 学习 公式.
doŋ ɦu *tsʐ̩ va* *ɦu zai* *koŋ sɛi*
classmate plan learn formula
'The classmate plans to learn the formula.'

b. 同学 允诺 学习 公式.
doŋ ɦu *joŋ nu* *ɦu zai* *koŋ sɛi*
classmate promise learn formula
'The classmate promises to learn the formula.'

c. 同学 允诺 老侄 学习 公式.
doŋ ɦu *joŋ nu* *lɜ dzai* *ɦu zai* *koŋ sɛi*
classmate promise nephew learn formula
'The classmate promises the nephew to learn the formula.'

[22] It should be noted that the English translation for the two embedding verbs *promise* and *encourage* does not seem to align with the syntactic analysis given here, which treats the NP intervening between the matrix and embedded verb as "embedded subject". In English, the intervening NP would have to be analyzed as object to the matrix clause in both cases. However, the results of the experiment, as presented in section 6.3 of this chapter, show that the Wenzhou speakers seemed to treat these intervening NPs as part of the embedded clause prosodically. The syntactic analysis of embedding structures in Chinese is notoriously difficult in the absence of overt morphological agreement marking, and different options for syntactic analyses of verbs comparable to those in the present experiment have been put forward for Standard Chinese (Huang 1987; Wei 1997). An in-depth analysis of the syntax of the embedding verbs in Wenzhou lies outside of the scope of this chapter.

d. 同学 鼓励 老侄 学习 公式.
doŋ ɦu *ku lɛi* *lɜ dzai ɦu zai* *koŋ sɛi*
classmate encourage nephew learn formula
'The classmate encourages the nephew to learn the formula.'

e. 同学 记得 老侄 学习 公式.
doŋ ɦu *tsʐ dɛi* *lɜ dzai ɦu zai* *koŋ sɛi*
classmate remember nephew learn formula
'The classmate remembers that the nephew learns the formula.'

The complex sentences were recorded with single word subject and object constituents, in order to keep the number of stimulus sentences within the experiment to a manageable size.

6.2.2 Speakers

Speakers were all between 20 and 29 years of age (mean age = 23;2) born and raised in the inner-city Lucheng district of Wenzhou. None of them reported to have lived outside of Wenzhou for a significant amount of time within the last 5 years, and they spoke the local dialect with their friends and family on a regular basis. All of them were fluent in Standard Chinese, but had no difficulty reading out aloud Chinese characters in their dialect. None reported any hearing or speech impediments. Nineteen speakers were recorded, 13 of whom were female. Of the 19 speakers, eight recorded the stimuli sentences once, and eleven twice. For those speakers with two recordings, the values were averaged over the recordings before statistical analysis.

6.2.3 Experimental procedure

Speakers were recorded in a sound-proofed recording studio in Wenzhou in individual sessions, and received a small payment for their participation. Each speaker was seated in front of a 13" monitor and given a Sennheiser pc130 headset. The experimenter ensured that the microphone of the headset was placed approximately 3 cm from the corner of the mouth of the subject. Via an external digitizer (UA-G1), the sound was recorded directly on the laptop (Acer Aspire 1810TZ) on which the stimuli were displayed to the subject.

The speakers were first informed about the recording procedure. They were instructed to read out phrases and sentences presented on the screen using Wenzhou dialect in a natural and clear manner. If they were unsure how to pronounce a word or phrase, they could skip to the next item; if they felt they

had made a mistake, they could go back and repeat the recording of the previous item. They were told that they could interrupt or abort the recording at any point.

The recording itself was done using a script in the computer program PRAAT (Boersma & Weenink 2001).[23] This script would present the stimulus sentences one by one, and record each stimulus individually after the speaker initiated the recording. Before the recording, all speakers completed a practice series with eight short phrases that were not part of the actual experiment. This was done in order to familiarize the speakers with the self-managed recording procedure. After completing the practice items, the speakers were asked to indicate whether they understood the recording procedure and were ready to start the experiment.

In order to ensure that speakers would not produce the sentences with narrow focus anywhere in the sentence, each sentence was presented together with a precursor question inducing broad focus over the entire sentence (你说什么啊？*ni kuɔ a ni a* – 'What are you saying?'), and the speakers were asked to read out both this question and the target sentence as an answer, as if enacting a dialogue. Furthermore, the target question-answer pairs alternated in the stimulus list with question-answer pairs from another experiment, in a manner that no speaker would see two of the target question-answer pairs from this experiment immediately in a row. This was done to ensure that, in spite of the repetitive nature of the target sentences, the speakers would not interpret the constituents as "given" in the discourse.

6.2.4 Data analysis

After the recording, all utterances were auditorily checked, and if found to contain mistakes and hesitations, excluded from further analysis. The remaining sentences were manually divided into words, using acoustic and visual inspection of the sound wave and spectrogram. A PRAAT script determined the F_0 peak within each disyllabic target word (defined as the F_0 maximum within that word), each F_0 valley between two successive F_0 peaks (defined as the F_0 minimum between two successive peaks) and at the beginning and end of the sentence, and recorded their positions and F_0 values[24]. Before F_0 extraction, the measurements were checked for octave jumps and tracking anomalies, and manually corrected where necessary.

[23] The script used for presenting and recording the stimuli was written by Jos Pacilly, and slightly modified by the author.

[24] The script used for measuring the files was written by Jos Pacilly.

Figure 6.1 shows an example sentence, with the bottom tier indicating word boundaries and the top tier indicating the automatically calculated locations of the F_0 peaks and valleys.

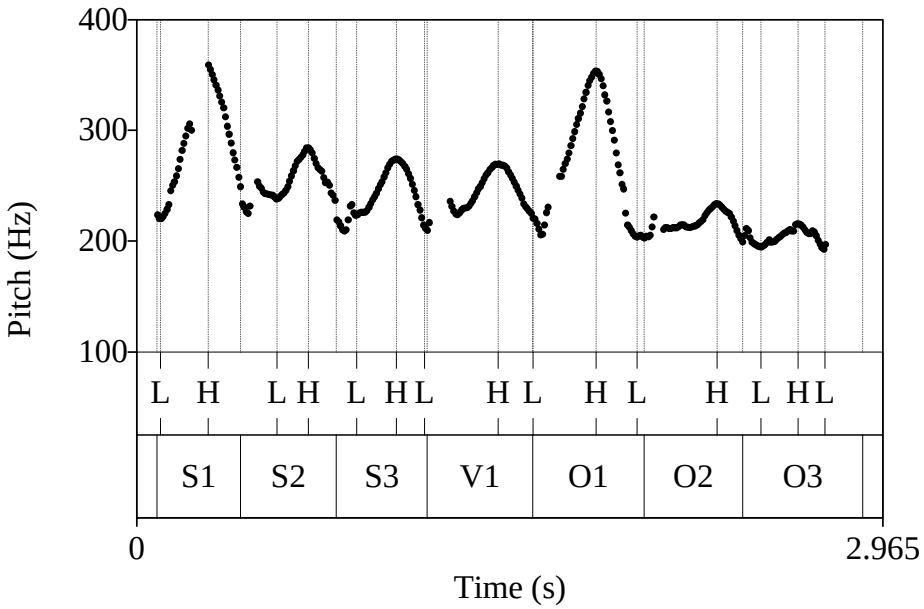


Figure 6.1: Example of a sentence with three subject and three object words, uttered by a female speaker. Dotted lines indicate the location of word boundaries and the automatically determined measurement points for F_0 peaks (H) and valleys (L).

The so determined F_0 values for the peaks (H) and valleys (L) were coded according to the constituent that they occurred on. For example, in Figure 6.1, H within S1 represents the first F_0 peak within the subject constituent, while H within O1 corresponds to the first peak of the object constituent. The so obtained measurements were compared between different combinations of constituent lengths. For plotting after this point, the F_0 values were converted to semitones by applying the formula 12*ln(Hz/x)/ln(2), with x being the pitch floor, which

was set to 50 hertz (Hz) for the male speakers, and to 100 Hz for the female speakers. All statistical analyses were performed on the original hertz values.[25]

6.3 Results

6.3.1 F_0 scaling in SVO sentences

The first research question was concerned with the effect of the length of the sentence on the scaling of the F_0 values. First of all, the height of the initial F_0 peak was investigated specifically, because this measure has been most commonly used to investigate the issue of sentential pre-planning. In order to investigate F_0 scaling as a function of the total sentence length, it was pooled over all tested combinations of subject and object lengths, and counted just the words within the sentences. A by-subjects Repeated Measures (RM) ANOVA was conducted with the height of the first subject peak (S1) as dependent variable and WORD number (five levels) as factor.[26] Results showed that the height of the first subject peak was indeed significantly affected by the number of WORDS in the sentence [$F(2.6,18) = 21.17$, $p < 0.001$].

However, in our data, sentence length as a measure is directly related to constituent length, as longer sentences are composed of longer constituents. To investigate the hypothesis that the length of the constituents exerts an independent influence on the scaling of the F_0 peaks, the data was split into the different combinations of subject and object lengths. The first investigation focused on the height of the first object peak and its relation to subject and object length. A by-subjects RM ANOVA was conducted with the height of the first subject peak (S1) as dependent variable and SUBJECT and OBJECT length (three levels each) as factors. Results showed that the height of the first subject peak was significantly affected by the length of the SUBJECT [$F(1.8,14) = 16.79$, $p < 0.001$], but independent of the length of the OBJECT [$F(1.84,14) = 3.0$, $p = 0.071$, ns]. This means that only the number of words in the subject constituent, but not the number of words in the object constituent affected the scaling of the first subject peak, as evident in Figure 6.2.

[25] Because the variances in the two speaker subgroups male vs. female are not equal, gender was not included as a factor in the statistical design.

[26] All reported degrees of freedom have been Huyhn-Feldt corrected when the requirement of sphericity was not met.

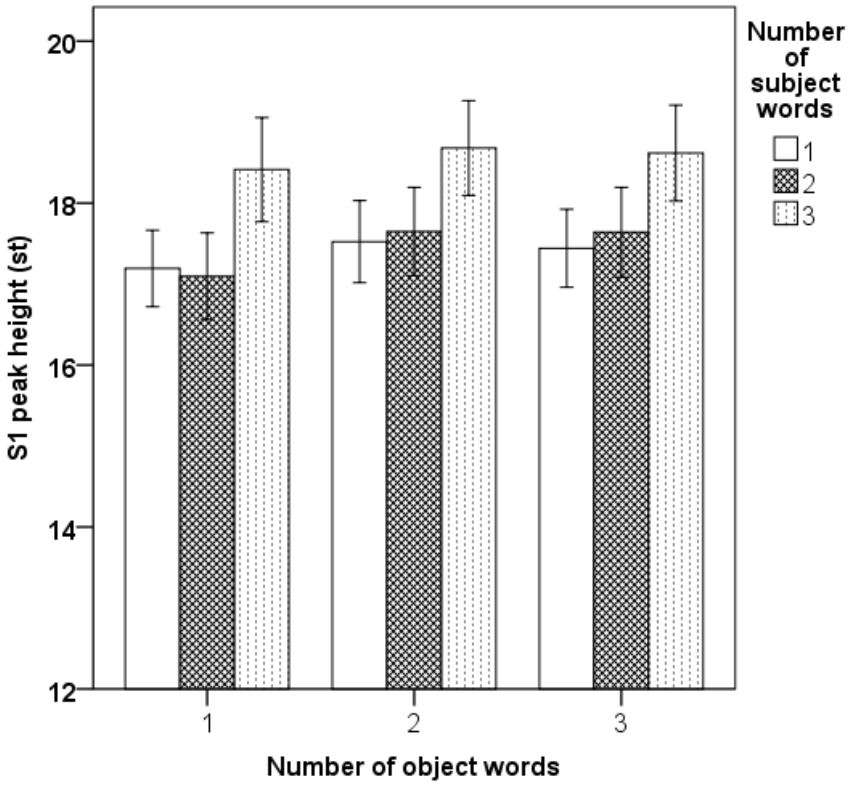


Figure 6.2: F_0 maxima (semitones) on the first subject peak, broken down by constituent length, averaged across speakers. T-bars = ±2 SE.

In order to investigate whether the opposite connection also holds, the height of the first object peak was set in relation to the length of the subject and object constituent respectively. A by-subjects RM ANOVA tested the dependence of the height of first object peak (O1) on the number of words in the SUBJECT and OBJECT constituent (three levels each). Results showed that, across all sentences, the height of the first object peak was dependent on the length of the OBJECT [$F(1.32,14) = 8.11$, $p < 0.01$], but independent of the length of the SUBJECT [$F(1.44,14) = 3.03$, $p = 0.084$, ns]. The results are illustrated in Figure 6.3. Conjointly, these results show that the effect of constituent length on the scaling of the initial peak of that constituent is independent of the length of the other constituent.

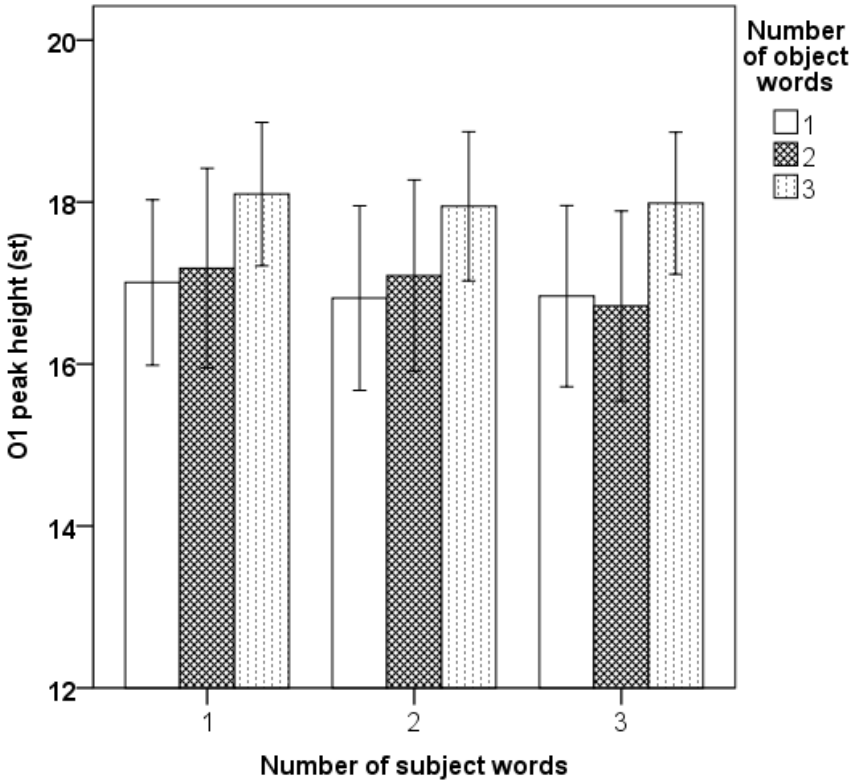


Figure 6.3: F_0 maxima (semitones) on the first object peak, broken down by constituent length, averaged across speakers. T-bars = ±2 SE.

Another interesting observation on the scaling of the first object peaks concerns the question of declination. As can be seen in Figure 6.3, the scaling of the first object peak is independent of the number of words in the subject constituent, even though more words in the subject constituent also mean that the first object peak comes at a later point in the sentence. Apparently, the height to which the first object peak is scaled is independent of its position within the sentence, and of how many words have already preceded it. This is an interesting observation, as it indicates that the scaling of the F_0 peaks in Wenzhou is unaffected by declination.

This finding is further corroborated by investigating the scaling of the F_0 minima across sentences of different lengths. The amount of lowering of the F_0 minima was obtained by subtracting the first from the last F_0 minimum value in each sentence, and then comparing the F_0 minimum ranges across the different sentence lengths (5 levels). A by-subjects RM ANOVA confirms that that there is no consistent increase in the amount of F_0 minima lowering with increased SENTENCE length [$F(3.33,18) = 1.72$, $p = 0.167$, ns]. These results suggest that our findings on the effect of constituent length on F_0 scaling are limited to the scaling of the F_0 peaks, without any declination effect on F_0 maxima or minima.

6.3.2 F_0 scaling in complex sentences

The second main research question concerns the sentential F_0 scaling properties of complex sentences. A comparison of different sentences with similar linear surface order, but different syntactic properties was intended to shed light on this question. First of all, Figure 6.4 illustrates the general tendency that is observed in all complex sentences: The syntactically embedded structures are also embedded in the F_0 scaling, which is lowered on both the F_0 peaks and F_0 valleys after the matrix verb.

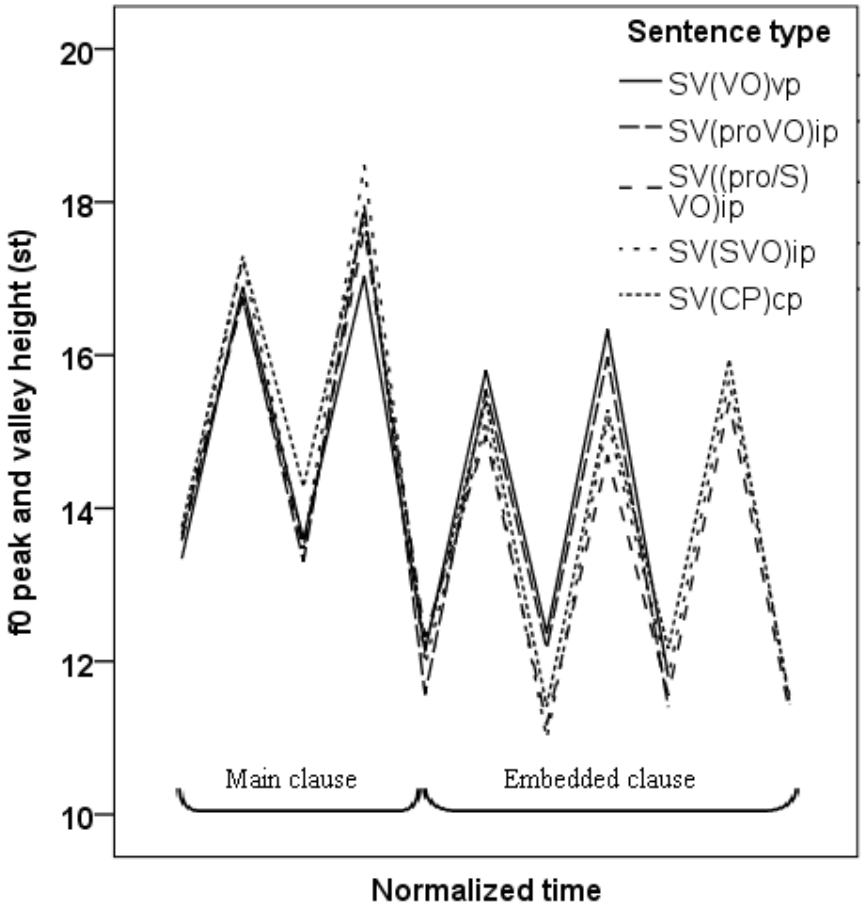


Figure 6.4: Results of F_0 maxima and F_0 minima (semitones) as a function of sequential position of constituents (normalized time), averaged across speakers, for complex sentences with four and five words.

In order to investigate the differences of the F_0 scaling between the sentence types in greater detail, the F_0 difference between the averaged peaks in the matrix clauses and the peaks in the embedded clauses was computed. Secondly, this difference, which represents the average magnitude of drop, was compared across the different sentence types. A by-subjects RM ANOVA showed that the magnitude of F_0 drop in the peaks differed significantly between the five

sentence types [$F(3.14,12) = 3.73$, $p < 0.05$]. Post-hoc analysis showed that between the five sentence types, the magnitude of drop for the $SV(VO)_{VP}$-sentences was significantly different from that of all other sentence types ($SV(PROVO)_{IP}$, $SV((PRO/S)VO)_{IP}$, $SV(SVO)_{IP}$, $SV(SVO)_{CP}$), but there was no significant difference between the other four sentence types. The results are graphically represented in Figure 6.5.

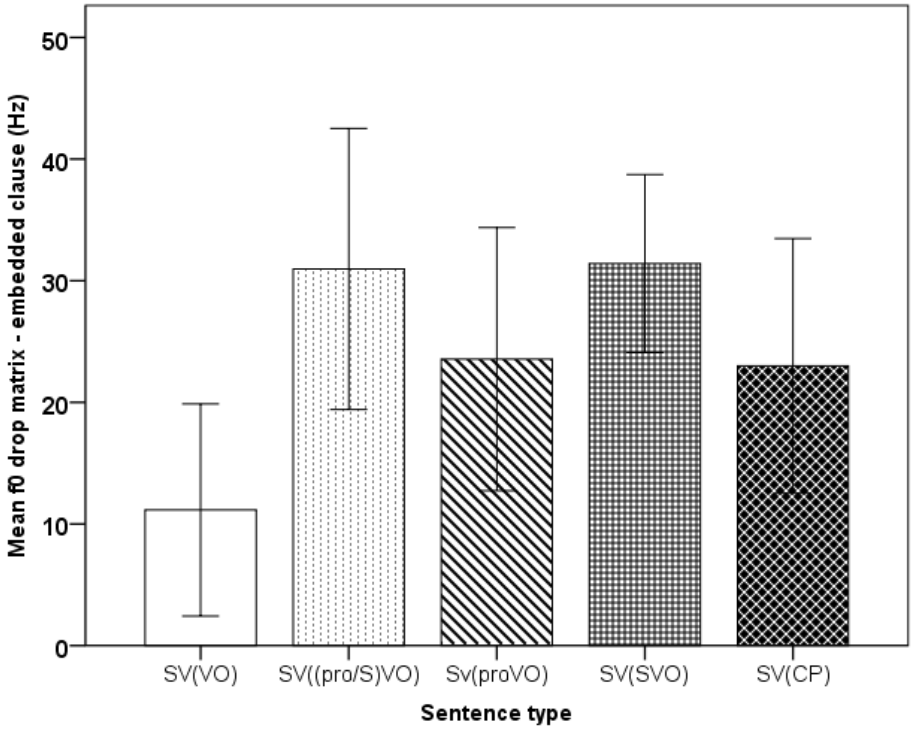


Figure 6.5: F_0 maxima drop (semitones) between matrix and embedded clause, averaged across speakers, for complex sentences with four and five words. T-bars = ±2 SE.

A similar result is obtained when comparing only the scaling of the last peak of the main clause (matrix verb) and the first peak of the embedded clause (embedded verb for $SV(VO)_{VP}$ and $SV(PROVO)_{IP}$, and embedded subject for $SV((PRO/S)VO)_{IP}$, $SV(SVO)_{IP}$, $SV(SVO)_{CP}$), [$F(3.54,12) = 4.04$, $p < 0.01$]. Post-hoc analysis again showed a significant difference between $SV(VO)_{VP}$ and the

other sentence types, but no difference among the other four sentence types ($SV(PROVO)_{IP}$, $SV((PRO/S)VO)_{IP}$, $SV(SVO)_{IP}$, $SV(SVO)_{CP}$. This confirms that the F_0 drop is indeed located at the boundary between the matrix and the embedded clause.

6.4 Discussion

This chapter aimed to shed light on two research questions concerning F_0 scaling in Wenzhou Chinese:

(i) Length effect: How does the length of the constituents determine the scaling of the initial peaks within these constituents, and (how) does it interact with the scaling of the F_0 valleys?

(ii) Complexity effect: In sentences with similar surface order and length that differ in underlying syntactic complexity, is the difference in complexity reflected in the F_0 scaling?

In the following paragraphs, the findings of the experiment will be connected to the respective research questions, and discussed with reference to previous research.

6.4.1 F_0 scaling as a function of constituent length

With regard to the first research question, this chapter first examined the connection between sentence length, constituent length, and the scaling of F_0 peaks. Results of our data showed an effect of sentence length on the scaling of the first subject peak. Further investigation, however, shows that the height of the initial F_0 peak is dependent on the length of the subject constituent, regardless of the length of the object constituent. In other words, the observed effect of sentence length is actually an effect of subject length, obscured by conflating the individual contribution of the length of the subject and object constituent. Furthermore, a consistent effect of object length on the scaling of the first object peak could be observed, which is independent of subject length. This finding lends support to the view that the speakers pre-plan the F_0 scaling of sentential constituents one at a time, and that when a new constituent starts, only the length of that constituent is taken into account in the pre-planning of the F_0 height of the initial peak. Preplanning in Wenzhou Chinese therefore is neither strictly “local” nor strictly “global”.

It is worth noting that independent of object F_0 peak scaling, this chapter also found a consistent resetting of the first object peak, i.e. a greater F_0 height on the first object peak than on the preceding verb, regardless of object length.

This resetting is reminiscent of what has been described for e.g. Japanese (Nagahara 1994; Selkirk & Tateishi 1991; Sugahara 2003). What differentiates Wenzhou Chinese from Japanese-type languages is the lack of rephrasing due to the length of the syntactic constituent. In Japanese, it has been observed that speakers opt to vary their prosodic phrasing so as to chunk an utterance into prosodic constituents of similar length and weight, rather than adhere to the syntactic boundaries (Hayashi 2004). Such a weight effect in prosodic phrasing has also been observed in some Romance languages (D'Imperio et al. 2005), and in English (Breen et al. 2011; Watson & Gibson 2004).

This weight effect has also been mentioned in an earlier investigation of Wenzhou Chinese (Chen 2000), but the current data set as presented in this chapter did not lend further supporting evidence. On the contrary, the data reported here shows that for the young Wenzhou speakers, syntactic and semantic coherence is the most important factor for prosodic phrasing. This is reflected in the consistency of the location of the F_0 reset, which occurs on the first object peak (O1) in all conditions. Apparently, the most important information that the speakers want to signal to the hearers is the beginning of a new sentential constituent, and not the approximate length of the upcoming constituent or the sentence as a whole. At the same time, by adjusting the height of the F_0 peaks in accordance to the number of words within the constituent, the speakers still include some information on the size of the upcoming material, while preserving the information of syntactic coherence.

A related issue concerning the scaling of F_0 peaks in Wenzhou Chinese is F_0 declination, i.e. continuous F_0 lowering across sentences. In most previous reports on downstep in tone languages including Mandarin (Wang & Xu 2011), the effects of downstep of F_0 peaks and declination of the F_0 valleys have been described as cumulative (see also Laniran & Clements 2003 for Yoruba). In Wenzhou, it seems that speakers reset their F_0 height to a stable level at the beginning of a new major syntactic constituent, and that this level is affected very little by sentence-internal overall declination of F_0 peak values (cf. Figure 6.3). Similarly, the declination of the baseline values (i.e. low tone targets or F_0 valleys) happens at a stable magnitude across sentences of different lengths. It is therefore independent of the F_0 peak reset discussed earlier, which occurs at a specific point in the syntactic structure, and it is also independent of sentence length.

6.4.2 F_0 scaling as a function of syntactic complexity

In addition to simple SVO sentences, the current chapter also investigated F_0 scaling of embedded clauses within a sentence. Specifically, it addressed the extent to which syntactic complexity is reflected in the intonational scaling of F_0 peaks in complex sentences in Wenzhou Chinese. The data show that the syntactically embedded sentences were also prosodically "embedded" into their matrix clauses, in the sense that both the F_0 peaks and valleys were scaled relatively lower in the embedded clause, compared to the matrix clause. On the surface, this non-local lowering of F_0 targets appears similar to what has been described as "downdrift" or "intonation register" (Yip 1993): A phonological trigger causes all subsequent tones within a prosodic constituent to be realized on an overall lower level. However, in the current study, the trigger for this overall lowering of F_0 range is not phonological (for example a low tone), but structural. The lowering occurs at the juncture of the syntactically embedded clause and its main clause.

Comparing different types of embedded clauses, however, it can be seen that speakers adjust their F_0 scaling to the syntactic complexity of the embedded clause only to a certain extent, even if its surface realization is similar to syntactically less complex structures. Specifically, the matrix subject and verb are realized with significantly higher F_0 peaks when preceding a clausal complement of any type (IP/ CP), than when preceding a VP complement, even when the IP complement has no overtly realized embedded subject. These findings can be interpreted to suggest that, by scaling the matrix structure higher, the speakers anticipate the complexity of the embedded structure and scale accordingly even when certain elements are not overtly realized. At the same time, it is interesting that the scaling of IP complements and CP complements is not very much different from one another.

The findings for complex clauses again underline the important role of syntactic structure for the implementation of F_0 targets in Wenzhou. It appears that it is an important function of F_0 scaling to signal the syntactic relationship between constituents in the sentence, be it clausal embedding (signaled by downshifting the entire F_0 range) or the beginning of a new phrasal constituent (signaled by partial resetting of the F_0 ceiling).

6.4.3 Possibilities for further research

As has been shown in section 6.3.2, the embedded SVO clauses have a similar internal F_0 scaling as the simple SVO clauses, in the sense that the F_0 height of the embedded object is also higher than that of the embedded verb. The

beginning of a new phrasal constituent within an embedded clause is therefore also marked with a reset of the F_0 ceiling, within the limits of the overall downshift of the F_0 range due to the syntactic embedding. It can be assumed that the clause-internal F_0 scaling in embedded clauses would be similar in cases with more than one subject and object word, so that the first peak of the subject or object constituent would be scaled higher dependent on the words within that constituent, but still lower than the matrix constituents. With the materials from the experiment reported in this chapter, it would be a straightforward possibility to test this prediction experimentally.

Another interesting possible follow-up research would be to test the influence of other factors that influence the F_0 scaling in sentences, such as focus. From research on Mandarin, it is known that focus has the effect of expanding the F_0 range and consequently raising the F_0 maximum of F_0 peaks, but it is still a matter of debate whether this expansion occurs globally over entire sentences (Shih 1988), or is locally concentrated on the focused constituent (Wang & Xu 2011). Previous research has suggested that prosody and focus may have similar effects on the surface (for example, both may result in a raising of the F_0 peak reference line), but that the two effects are phonetically distinguishable and therefore should not be treated as two sides of the same coin, but as independent factors (Chen 2004; Ishihara 2011; Wang & Xu 2011). For Wenzhou Chinese, it would be interesting to investigate the influence of focus on the F_0 scaling effects reported in this chapter, and to compare the observed similarities and differences to the findings in other languages.

6.5 Conclusion

This chapter investigates the relationship between sentence length, constituent length, syntactic complexity, and F_0 scaling in sentences in Wenzhou Chinese. By varying the length of the constituents in SVO sentences independently of each other, the scaling of the sentence-initial F_0 peak in Wenzhou is shown to be determined by the number of words within the subject constituent, regardless of the length of the object constituent. Likewise, the scaling of the first peak within the object constituent in SVO sentences is solely dependent on the number of words within the object constituent, regardless of the length of the subject.

Within the SVO sentences, the speakers displayed a stable tendency to "reset" the F_0 peak at the beginning of the object constituent. This finding speaks for a syntactically derived prosodic structure, which governs F_0 scaling across sentences in Wenzhou. At the same time, the height of this F_0 reset is resistant to

the influence of any general effect of sentential lowering (declination). This suggests that downstep and declination in Wenzhou are not cumulative effects, but that downstep is able to temporarily suspend topline declination.

In complex sentences with embedded VP, IP, or CP complements, the entire F_0 range of the embedded F_0 peaks and valleys was lowered and compressed. At the same time, the relative scaling of the constituents within the embedded clause with respect to each other was similar to that in simple SVO sentences. This leads to the possibility that the lowering of embedded clauses and the reset at phrasal constituent edges are separate processes that occur independently of each other. The reset on the object that occurs within a clausal constituent is constrained by the overall scaling of that constituent in the sentence.

At a different level, a significant prosodic difference between VP complements and clausal complements can be observed, with clausal complements showing a steeper drop of F_0 peaks between the matrix clause and an embedded complex complement, compared to an embedded simple VP complement. In that sense, the speakers are able to mark complex syntactic configurations intonationally, and seem to attempt to give more "room" to the intonational realization of more complex structures. Such a distinction, however, is not present for F_0 scaling of IP vs. CP complements.

Chapter 7

The marking of information focus in Wenzhou Chinese

7.1 Introduction

7.1.1 Effects of information focus

7.1.1.1 Lexical effects

One of the most common notions in the research on information structure is the notion of focus. Broadly speaking, the focus of a sentence is the constituent that is most informative or most important in a sentence. Such importance or informativeness can be brought about in different ways. One of the most commonly investigated types of focus has been called "information(al) focus" (É. Kiss 1998; Gundel & Fretheim 2006; Jackendoff 1972), "presentational focus" (Gussenhoven 2007), or "narrow focus" (Féry & Kügler 2008). It is usually elicited by drawing up a question-answer-pair, and the idea is that the constituent in the answer that corresponds to the wh-element in the question is in focus, an idea that is often credited to Halliday 1967.

An important consequence of having a focus in a sentence or utterance is that it affects the cognitive status of the person or thing that is within the focus domain. In the semantic notion of a *common ground* that is shared between the hearer and the speaker in a discourse, the common ground contains (shared) knowledge of propositions and referents within the realm of the discourse (see e.g. Krifka 2007 for a review of the terminology). The status of a referent in the common ground can be indicated by the use of specific linguistic expressions, such as pronouns and (definite/indefinite) articles. It has been proposed that the linguistic means that speakers use to refer to things or persons in a discourse are hierarchically ranked, and that DP-structures such as demonstrative pronouns, definite articles, and indefinite articles represent a decreasing order of activation within the discourse (Gundel et al. 1993).

Crucially, a referent that is referred to with e.g. a demonstrative pronoun has a higher likelihood to be given/familiar in the discourse than a referent that

is referred to with an indefinite article. In that sense, the focus/givenness structure of a sentence interacts with the cognitive status of the referents, and consequently influences the types of referential expressions that are used by speakers to convey information to hearers. In the *Givenness Hierarchy* proposed by Gundel et al. 1993, a referent that is *in focus* is endowed with the highest state of activation within the discourse, and can consequently be referred to with e.g. a zero or unstressed pronominal, while the speaker can still make the assumption that the hearer will be able to identify the referent correctly.

7.1.1.2 Phonetic effects

Another important effect of information focus is its influence on the intonation of sentences and utterances. In intonation languages such as English or German, focus on a specific constituent in a sentence changes the implementation of the accent that is associated with that constituent, both vertically (the accent reaches a higher F_0 maximum) and horizontally (the accent-bearing unit is increased in duration). Additionally, the accents that follow the focused constituent within the same prosodic domain are often demoted in prominence by virtue of lowering (Féry & Kügler 2008). In pitch-accent languages such as (Tokyo) Japanese, focus has been found to be marked in a similar way: the F_0 peak on the focused constituent is raised via manipulation of pitch range, and the post-focal accents are compressed in vertical range (Ishihara 2007, 2011).

Apart from the specific acoustic parameters that are affected by focus, a second important research question is concerned with the marking of focus domains of different sizes. For English (Bishop 2010; Breen et al. 2010; Eady et al. 1986), Dutch (Hanssen et al. 2008), and German (Baumann et al. 2007; Baumann et al. 2006; Kügler 2008), it has been experimentally shown that broad focus, more specifically a focus domain that involves the entire sentence, is marked in a different way (namely with less expanded acoustic parameters) than a narrow focus which includes only one of the constituents in the focus domain. While for example in German, both broad and narrow focus are marked with pitch accents and lengthening, the precise phonetic realization varies between the two types of focus, such that a narrow focus receives more lengthening and greater F_0 expansion on the pitch accent than broad focus. This aligns with findings for languages such as Korean, which also show a marked difference in the realization of broad (VP) vs. narrow (object) focus (Jun & Kim 2007).

7.1.2 Phonetic effects of information focus in Standard Chinese

Most experimental research on the phonetic correlates of focus realization in Chinese has been conducted for Mandarin dialects of Chinese. Challenging the common assumption that tone languages do only use intonation in a limited manner, since their "intonation" is already needed for lexical disambiguation, it has been found by several studies that speakers of Mandarin employ similar phonetic cues for focus marking as accent-type and intonation languages. Particularly, focused constituents were found to be realized with expanded pitch ranges, lengthening of the focused constituent, and compressed pitch ranges post-focally.

Specifically, for wh-induced focus as in the present experiment, Xu (1999) determined the precise effect on constituents with different citation tones. He reports that under focus, the F_0 maxima of the high level, rising, and falling tones were raised, while the F_0 minima of the low/dipping, rising, and falling tones were lowered (see also Kabagema-Bilan et al. 2011 for similar results). Therefore, expansion of the pitch range due to focus appears to target both directions, which sets it apart from the prosodic marking of (new) topics, where both F_0 maxima and F_0 minima are raised (Wang & Xu 2006).

In addition to pitch range expansion, lengthening of the focused constituents has been determined as another stable cue (Jin 1996; Pan et al. 2005; Wang & Xu 2006). Furthermore, post-focal tones have been found to be realized with a significantly lowered F_0 compared to control conditions (Xu et al. 2004), and together with the also lowered intensity on post-focus constituents, this may serve as an additional cue for listeners to determine the focus position within a sentence in perception tests (Chen et al. 2009; Xu et al. 2012).

However, the above described focus effects are not found in all sentence positions: Jin 1996) reports that sentence-final lengthening obscures the lengthening effect of focus on sentence-final constituents in Mandarin, so that a focused constituent in sentence-final position is prosodically indistinguishable from non-focus control condition. At the same time, F_0 expansion in sentence-final position is also remarkably lower than on focused constituents in sentence-initial position. This results in the F_0 range of broad focus and narrow focus condition in sentence-final position being not significantly different from one another.

These findings can be partially attributed to the fact that F_0 tends to decline over the course of an utterance. However, as Xu (1999) reports for his investigation of different tone combinations in sentences, this downtrend is tone-dependent. More specifically, he reports that in a sentence consisting of high

tones only, the difference in F_0 height between successive syllables is very small. In contrast, all other tones induce declination, in a way that the declination increases with the number of non-high tones present in a sentence. Therefore, in 'natural' sentences consisting of more than just high level-toned syllables, declination can be expected, and consequently an early focus in the sentence should result in more pitch range expansion than a late focus.

7.1.3 Phonetic effects of focus in other dialects of Chinese

Most recently, the effects of wh-induced information focus have also been investigated for other dialects of Chinese. For Shanghai Chinese, F_0 range expansion and lengthening on the focused constituent, as well as lowered F_0 values in the post-focal tones have been reported (Chen 2009). This pattern is comparable to the focus effects that have been found in Standard Chinese, but it preserves the dialect-specific tone sandhi characteristics and inherent length differences between syllables.

For other dialects such as Taiwanese, it has been found that the focus effect is dependent on the specific tone of the focused syllable. In Pan 2007b, it was found that F_0 range expansion on focused constituents could only be reported for some participants and some specific tones, namely the contour tones, whereas level tones tended to remain unaffected by focus in their F_0 implementation. In contrast, duration proved to be a very stable cue for focus marking, even though the amount of lengthening was dependent on sentence position, such that syllables in pre-final sentence position showed the least amount of focus-induced lengthening.

This finding has recently been challenged in a cross-dialectal comparison that included Taiwanese, Taiwan Mandarin and Beijing Mandarin (Chen et al. 2009; Xu et al. 2012). The authors found that both monolingual and bilingual speakers of Taiwanese had very little alteration in the F_0 contours of identical sentences over different focus conditions, and also did not use lengthening as a stable cue for focus position within a sentence. Even though focused syllables tended to be longer, post- and pre-focal syllables were also lengthened by the speakers as soon as there was a focus somewhere in the sentence, which obscured the marking effect of lengthening on the focused syllable(s).

Interestingly, the focus realization by both monolingual and bilingual (Taiwanese and Taiwan Mandarin) speakers of Taiwan Mandarin patterned with the observations made for Taiwanese, even though Taiwan Mandarin and Beijing Mandarin are more closely related. The authors conclude that focus

marking strategies are prone to contact-induced change and therefore, the prosodic cues for focus marking have been lost in Taiwan Mandarin. This suggestion is confirmed by another study, that finds that Taiwan Mandarin speakers are unable to perceptually distinguish different focus positions within a sentence (Pan et al. 2005). This suggests that, not only is focus marking different across different dialects of China, but also highly influenced by the amount of exposure that the speakers will have to different dialects and their (different) focus marking strategies.

7.1.4 Experimental approaches

The current chapter reports the results of two experiments which individually tested the influence of focus on the two different parameters outlined above. In the first experiment, a picture elicitation method is used in order to investigate the focus-marking strategies that the speakers employ in a more "naturalistic" experimental setup. The experimental prompt consisted of acoustic stimulation and pictures only, and the research question concerns the specific types of linguistic expressions that speakers use to refer to new, given, or focused discourse referents. By comparing the length and specificity of the grammatical expressions used to describe the referents in different discourse situations, this experiment intends to shed light on the discourse strategies employed by the speakers.

In order to allow for a direct comparison with the published findings concerning the phonetic effects of focus in Chinese, the second experiment uses written dialogues in the form of question-answer pairs. Similar scenarios as in the first experiment are used, but this time the speakers are asked to read out the mini-dialogues in Wenzhou dialect, as prompted by the written version of the dialogues in Standard Chinese characters. By adjusting the target sentences to the Wenzhou vernacular (for example, including a commonly used aspect marker), the sentences are still naturalistic, but the speakers will be more limited in their realization of the target sentences. This allows for a detailed acoustic analysis of the realizations along the phonetic parameters that have been established for other dialects of Chinese. In this way, it can be assessed whether the phonetic markings of focus that are used by Wenzhou speakers are different from the phonetic markings used by speakers of other Chinese dialects.

7.2 Experiment 1: Lexical realization of referents

7.2.1 Stimuli

In order to investigate the referential expressions that are employed by the speakers in different discourse situations, a picture description paradigm was used in the first experiment. The stimulus pictures that were used were originally developed for the research group SFB 632 in Berlin and Potsdam (Skopeteas et al. 2006) for cross-linguistic investigation of information-structural categories.[27] Pictures were selected on the criterion that the action described in the picture can be expressed with a simple transitive verb (*hit, kick, push, pull*). Concerning the referents in the pictures, both animate (human, non-human) and inanimate referents were included in the picture selection. For a full list of stimulus pictures used in the experiment, the reader is referred to appendix 7.1.

The pictures were paired with context questions which prompted a certain focus structure. The context questions were recorded prior to the actual experiment by a male speaker who was in the same age group as the recording subjects. The speaker was presented with individual pictures that were paired with the intended context question and target answer sentence in English, and was asked to translate both the question and the answer in his head, and then pronounce both in Wenzhou dialect. English rather than Standard Chinese was used as the elicitation language in order to minimize the influence of Standard Chinese, and ensure a naturalistic Wenzhou rendition rather than a word-by-word translation of the Standard Chinese sentences. By eliciting both the intended question and the answer, the experimenter could check the prosodic realization of both sentences to ensure that the speaker had rendered the focus structure correctly.

If a question was incomplete, contained an unusual word order or passive voice, or sounded unnatural to the experimenter, the speaker was asked to repeat the question, and encouraged to think of "another way to say it". This was intended to ensure that the questions were in SVO word order and in the active voice. Two exceptions occurred: the speaker realized two examples with SOV word order, one with an inanimate subject referent in subject focus ("the bike hits the woman"), and one with two identical referents ("the man hits the other man") in object focus. The speaker judged the sentences to be more natural with this word order in the specific focus conditions.

[27] Website of picture materials: http://www.sfb632.uni-potsdam.de/~d2/materials.php#refer; last accessed 6-8-2012.

Three focus structures were investigated in the first experiment: subject focus, VP focus, and object focus. In both the VP and the object focus context, the subject referent was included in the question and therefore counted as "given" in the answer target sentence. Conversely, in the subject focus condition, the object referent was "given" in the question. The speakers of the experiment were encouraged to answer the context questions with full target sentences, and to realize both the focused and the given referents in all conditions. Figure 7.1 represents one of the stimulus pictures, and example (1) illustrates a transcription of (one possible realization of) the intended target sentence and the recorded context questions that were presented together with the picture.[28]

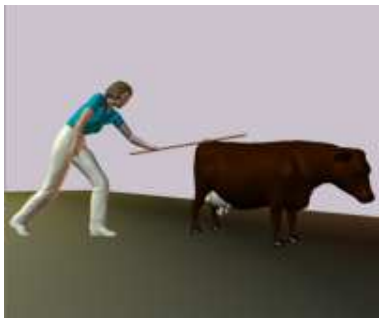


Figure 7.1: Example stimulus picture from the QUIS corpus (Skopeteas et al. 2006: 79, item 2 picture 2). Reprinted with permission of the SFB 632/ Malte Zimmermann.

(1) **Intended target sentence**:

a.	*ʔi*	*kai*	*nø*	*zẓ ta*	*tiɛ*	*ʔi*	*noŋ*	*ŋau*
	NUM	CL	woman	ASP	hit	NUM	CL	cow
	'The woman is hitting the cow.'							
	Subject focus context question:							
b.	*a ni*	*naŋ*	*(a)*	*zẓ ta*	*tiɛ*	*ki*	*noŋ*	*ŋau?*
	which	person	Q	ASP	hit	this	CL	cow
	'Who is hitting the cow?'							

[28] Because there was no prompt in Chinese characters in this experiment, the example sentences will only be transcribed in a broad transcription here and in the appendix.

VP focus context question:

c. *kai* *nø* *zɻ ta* *tɕø* *(a) ni?*
CL female ASP do what
'What is the woman doing?'

Object focus context question:

d. *kai* *nø* *ɻ ta* *tiɛ* *a ni* *mɵ zɻ?*
CL female ASP hit what thing
'What is the woman hitting?'

7.2.2 Speakers

The subjects of this experiment were eight speakers (three male, five female) between 18 and 20 years of age. They were high school graduates of the same high school in central Wenzhou and all born and raised in the central district of Lucheng Wenzhou. None of them reported to have lived outside of Wenzhou for a significant amount of time within the last five years, and all of them considered themselves fluent speakers of the Wenzhou dialect. They were also fluent speakers of Standard Mandarin, which they learned in school and used in conversations on a daily basis.

7.2.3 Experimental procedure

The recordings were made in a quiet recording studio in the TV and radio station in Wenzhou on an M-Audio Microtrack II portable digital recorder in wav-format (44.1 kHz, 16bits mono). The speakers were given a Sennheiser pc130 headset, and the position of the microphone was adjusted by the experimenter to ensure it was about 3 cm away from the corner of the mouth and outside of the immediate direction of exhalation.

Each speakers was seated at a table with about 50 cm distance from a laptop screen (ACER TravelMate 280XCi), on which the stimulus pictures were presented using E-Prime© software. All speakers confirmed they could see the pictures properly. The stimulus pictures were automatically randomized for every speaker and every trial by E-Prime, and presented in an individual fashion, with the speaker determining the pace of succession. Alongside with every picture, the speaker heard the pre-recorded stimulus question (see section 7.2.1) over SONY loudspeakers that were attached to the laptop. The speakers also confirmed that they could hear the questions clearly in a practice session before the actual experiment began.

Before the start of the experiment, the speakers saw a welcome text in Standard Chinese that informed them about the task they were asked to perform

(see appendix 7.2). They were told that they would see pictures and hear a question, and were to answer the question in a clear and natural way and with a complete sentence, using the information provided in the picture. Next, they were presented with a practice picture and two context questions (subject and object focus), to prepare the speakers for the fact that they might see one picture multiple times, paired with different context questions each time. The questions were played acoustically over loudspeakers, and followed by an example of an incomplete answer (only focused NP) and a complete answer (entire sentence), in order to prompt the speakers to use complete sentences in their answers.

After they had seen the practice items, they were asked to confirm to the experimenter that they had understood the task and the difference between the complete and the incomplete practice answers, and then proceeded through the experiment in a self-paced manner. Each speaker was presented with the entire task three times in a row, and asked to take a short break in between repetitions. They received a small payment for their participation.

7.2.4 Data analysis

The choice to impose fewer restrictions on the speakers by giving them a picture description task, as opposed to a reading-aloud task, resulted in a wide range of variation both across and within speakers in terms of lexical and structural choices in the rendition of the target sentences. A comparison of the different structures across focus conditions yielded the observation that the speakers systematically varied the complexity of two aspects of the realization of the referents, namely the length and the definiteness of the DP structures they were denoted with. For each of the aspects, a classification scale was drawn up to categorize the observed structures. For length, the number of syllables within the NP plus preceding adjectival modifiers were counted, as illustrated in example (2). For definiteness, different kinds of anaphoric specificity were coded, as illustrated in example (3). In both examples, the respective syllables are given in bold.

(2) **Length: Number of syllables**

a.	One syllable:	*kai*	***nø***	*zẓ ta…*
		CL	female	ASP
		‘The woman is...’		
b.	Two syllables:	*kai*	***nø ki***	*zẓ ta…*
		CL	female	ASP
		‘The woman is...’		

c.	Three syllables:	*kai*	***nø***	***mai mai***		*zẓ ta…*
		CL	female	little.child		ASP
		‘The girl is...’				
d.	Four syllables:	*kai*	***sai***	***nø***	***mai mai***	*zẓ ta…*
		CL	small	female	little.child	ASP
		‘The little girl is...’				

(3) **Specificity: Determiner/numeral/classifier within the DP**

a.	Bare noun:				*nø*	*zẓ ta…*
					female	ASP
		‘(A/the) woman is...’				
b.	Classifier+noun:			***kai***	*nø*	*zẓ ta…*
				CL	female	ASP
		‘The woman is...’				
c.	Numeral+classifier+noun:	***ʑi***		***kai***	*nø*	*zẓ ta…*
		NUM		CL	female	ASP
		‘A woman is...’				
d.	Demonstrative+classifier+noun:	***ki***		***kai***	*nø*	*zẓ ta…*
		this		CL	female	ASP
		‘This woman is...’				
e.	Existential+classifier+noun:	***jau***		***kai***	*nø*	*zẓ ta…*
		there.is		CL	female	ASP
		‘There is a woman who is...’				
f.	other+classifier+noun:	***loŋ vai***		***kai***	*nø*	*zẓ ta…*
		the.other		CL	female	ASP
		‘The other woman is...’				
g.	other+numeral+classifier+noun:	***loŋ vai***	***ʑi***	***kai***	*nø*	*zẓ ta…*
		the.other	NUM	CL	female	ASP
		‘The other (one) woman is...’				
h.	Numeral+demonstrative+ classifier+noun:[29]	***ʑi***	***ki***	***kai***	*nø*	*zẓ ta…*
		NUM	this	CL	female	ASP
		‘This (one) woman is...’				

[29] As can be seen in the results section below, there are only four instances in 993 sentences with this structure in the entire experiment. It is unclear whether this is a legitimate structure, or should be counted as mispronunciation by the speakers. In Standard Chinese, the structure would not be well-formed.

The so-coded data was compared across the different focus conditions, to determine whether speakers have a preference to use a specific DP-type in a certain focus situation, and whether they systematically vary the syllable count in the constituents in correspondence with certain focus contexts.

7.2.5 Results

7.2.5.1 Syllable count within the NP

The pictures that were presented to the speakers all depicted actions that can be described with a monotransitive verb. Therefore, all the sentences recorded contained two referents, namely a subject referent and an object referent. The precursor questions were recorded in a way to ensure SVO word order, with the expectation that this would prompt speakers to realize their answers with this constituent order as well. Therefore, the subject referent always precedes the object referent in the answers.

Table 7.1 and Figure 7.2 show the results for subject referents, pooled across the three repetitions for all speakers of all 14 stimulus pictures, and split according to the three focus conditions. Depicted is the frequency with which speakers used the different NP lengths, as exemplified in (2). The numbers in the legend correspond to the number of syllables within the NP-constituent.

Table 7.1: Absolute and relative frequencies of length of subject constituent, crosstabulated by focus condition. Overall token results.

Length of subject in syllables	Focus on			Total
	Object	Subject	VP	
1	176 (52.9%)	92 (27.4%)	150 (46.2%)	418 (42.1%)
2	40 (12.0%)	66 (19.6%)	43 (13.2%)	149 (15.0%)
3	117 (35.1%)	168 (50.0%)	129 (39.7%)	414 (41.6%)
4	0 (.0%)	8 (2.4%)	3 (.9%)	11 (1.1%)
5	0 (.0%)	2 (.6%)	0 (.0%)	2 (.2%)
Total	333	336	325	994

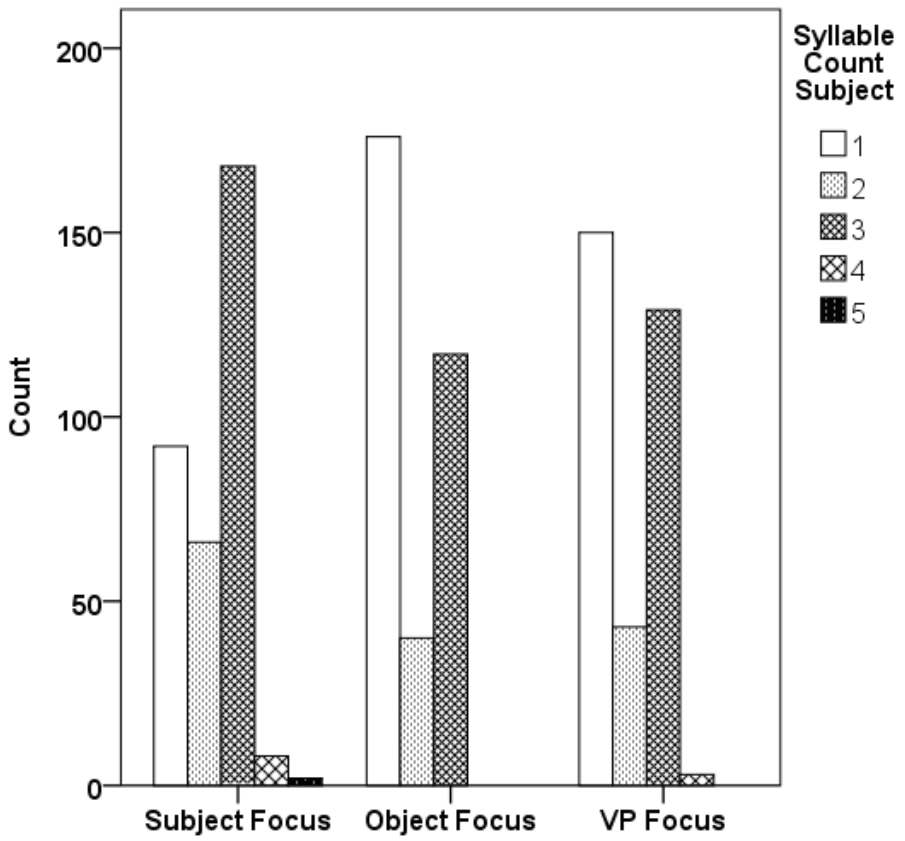


Figure 7.2: Count of instances of subject NPs with one to five syllables, broken down by focus condition.

As can be seen, there is a systematic interaction between the focus/givenness status of the referent and the number of syllables used to denote it. In object and VP focus condition, both of which have the subject referent given in the precursor question, the number of monosyllabic NPs (blank bar) is much higher than in subject focus condition, and conversely, the number of disyllabic (light grey bar) and trisyllabic (dark grey bar) realizations is lower. This corresponds to the expectation that a given referent is often demoted in prominence, which can be related to the length of its realization. A Pearson Chi-square statistical test of the results for the length of the subject constituent confirms a significant difference in length between the three focus conditions [$\chi^2(8) = 57.18$, $p < 0.001$].

The inverse picture can be observed for the object referent, as illustrated in Figure 7.3 and Table 7.2. The likelihood for an object NP to contain two rather than one syllable is much higher under object focus and VP focus than under subject focus. A Pearson Chi-square statistical test of the results for the length of the object constituent again confirms a significant difference in length between the three focus conditions [$\chi^2(10) = 41.75$, $p < 0.001$].

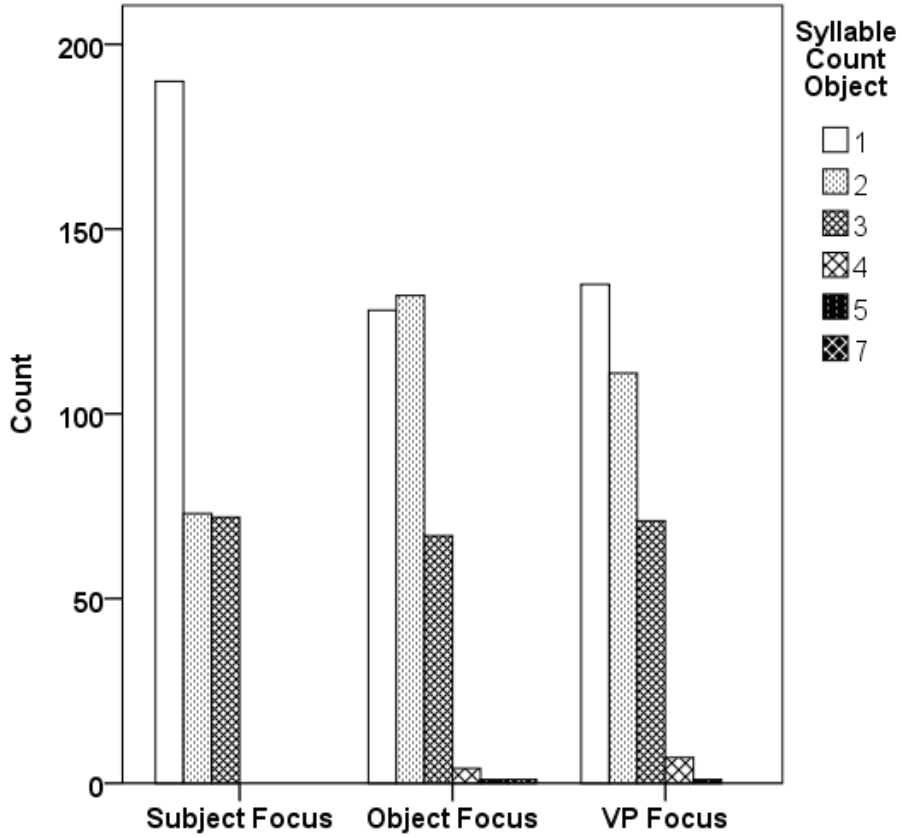


Figure 7.3: Count of instances of object NPs with 1-7 syllables, broken down by focus condition.

Table 7.2: Absolute and relative frequencies of length of object constituent, crosstabulated by focus condition. Overall token results.

Length of object in syllables	Focus on			Total
	Object	Subject	VP	
1	128 (38.4%)	190 (56.7%)	135 (41.5%)	453 (45.6%)
2	132 (39.6%)	73 (21.8%)	111 (34.2%)	316 (31.8%)
3	67 (20.1%)	72 (21.5%)	71 (21.8%)	210 (21.1%)
4	4 (1.2%)	0 (.0%)	7 (2.2%)	11 (1.1%)
5	1 (.3%)	0 (.0%)	1 (.3%)	2 (.2%)
7	1 (.3%)	0 (.0%)	0 (.0%)	1 (.3%)
Total	333	335	325	993

7.2.5.2 Definiteness of the DP

Figure 7.4 shows the distribution of the different DP types that were recorded in the experiment. The variants are coded in numbers, with the correspondences being as follows (compare (3) for examples and transcription):

(4) **Coding of DP specificity**
0 = Bare noun
1 = Classifier+noun
2 = Numeral+classifier+noun
3 = Demonstrative+classifier+noun
4 = Existential+classifier+noun
5 = “other”+classifier+noun
6 = “other”+numeral+classifier+noun
7 = Numeral+demonstrative+classifier+noun

As summarized in section 7.1.1.1, the assumption is that the definiteness of a referring expression is correlated with the familiarity/givenness of the respective referent within the discourse. Referents that are familiar or uniquely identifiable and thereby given in the discourse are expected to be realized with a demonstrative+noun or a definite structure, whereas unfamiliar referents which

are newly introduced into the discourse are more commonly realized with an indefinite structure (cf. Gundel et al. 1993).

For Wenzhou, it has been argued that the classifier+noun structure, as in (3b), encodes definiteness when the referent occurs in preverbal position. Conversely, the numeral+classifier+noun structure is commonly interpreted as indefinite, and may be either specific or unspecific (Cheng & Sybesma 1999, 2005; Li & Bisang 2012). Cheng & Sybesma (2005) argue that the indefinite numeral+classifier+noun structure only occurs in postverbal position, but as the results below show, this is not borne out by the results of the current experiment. An explanation for the divergent findings may be that the (postverbal) object position is the most common location for default focus (Xu, L. 2004). Therefore, in an analysis that does not take focus structure into account, it is more likely that the indefiniteness which is associated with focus is more often encountered in object position. For the current analysis, the following predictions concerning referent realization can be hypothesized.

(5)

a. **Givenness hierarchy** (after Gundel et al. 1993)
Decreasing familiarity/givenness from left to right):
Demonstrative > Definite structure > Indefinite structure

b. **Definiteness hierarchy in Wenzhou** (after Cheng & Sybesma 2005)
Decreasing definiteness from left to right:
Demonstrative+classifier+noun>Classifier+noun>
Numeral+classifier+noun

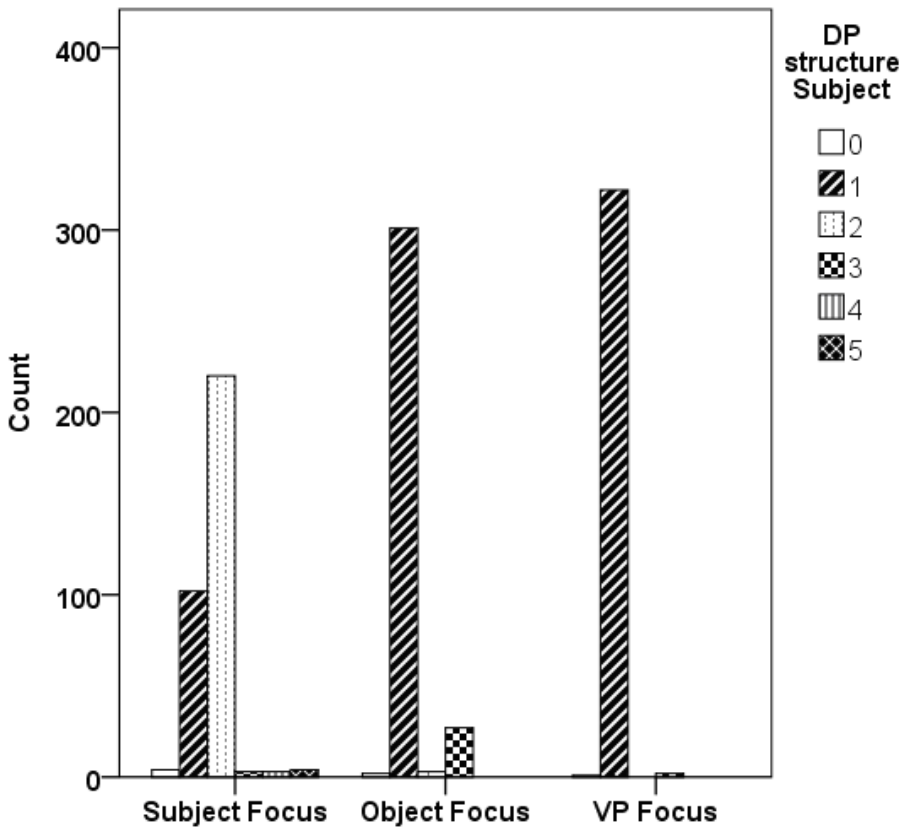


Figure 7.4: Count of instances of subject DPs with different degree of specificity, broken down by focus condition.

As can be seen in Figure 7.4 and Table 7.3, there is indeed an interaction between DP-specificity and focus condition. Under subject focus, the speakers prefer the indefinite numeral+classifier+noun structure (light grey bar) to denote the focused referent, whereas under VP and object focus with a given subject referent, they most often make use of the definite classifier+noun structure (diagonally striped bar) to denote the subject. A Pearson Chi-square statistical test of the results for the specificity of the subject constituent confirms a significant difference between the three focus conditions [$\chi^2(10) = 599.9$, $p < 0.001$].

Table 7.3: Absolute and relative frequencies of specificity of subject constituent, crosstabulated by focus condition. Overall token results.

Specificity of subject	Focus on			Total
	Object	Subject	VP	
0	2 (.6%)	4 (1.2%)	1 (.3%)	7 (.7%)
1	301 (90.4%)	102 (30.4%)	322 (99.1%)	725 (72.9%)
2	3 (.9%)	220 (65.5%)	0 (.0%)	223 (22.4%)
3	27 (8.1%)	3 (.9%)	2 (.6%)	32 (3.2%)
4	0 (.0%)	3 (.9%)	0 (.0%)	3 (.9%)
5	0 (.0%)	4 (1.2%)	0 (.0%)	4 (1.2%)
Total	333	336	325	994

Conversely, when the object is given under subject focus, the speakers predominantly use the definite demonstrative+classifier+noun structure (checked pattern bar) in to denote the object referent, whereas under object and VP focus, the indefinite numeral+classifier+noun structure (light grey bar) is most prevalent, as can be seen in Figure 7.5 and Table 7.4. A Pearson Chi-square test of the results for the specificity of the object constituent confirms a significant difference between the three focus conditions [$\chi^2(14) = 334.81$, $p < 0.001$].

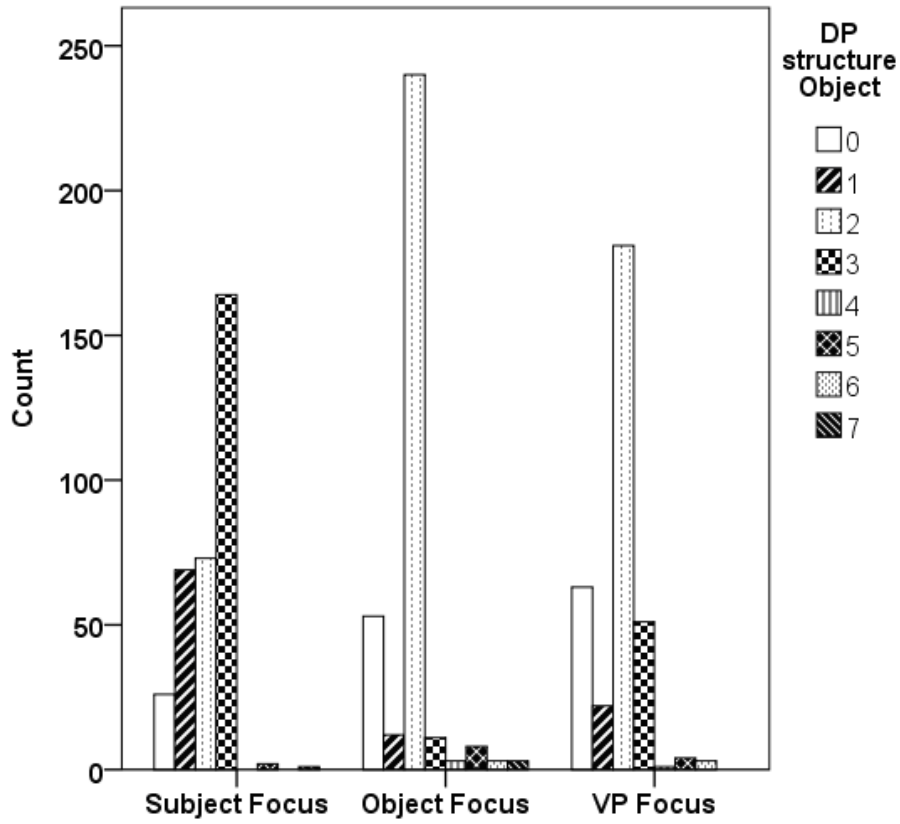


Figure 7.5: Count of instances of object DPs with different structures, broken down by focus condition.

Table 7.4: Absolute and relative frequencies of object constituent, crosstabulated by focus condition. Overall token results.

Specificity of object	Focus on			Total
	Object	Subject	VP	
0	53 (15.9%)	26 (7.8%)	63 (19.4%)	142 (14.3%)
1	12 (3.6%)	69 (20.6%)	22 (6.8%)	103 (10.4%)
2	240 (72.1%)	73 (21.8%)	181 (55.7%)	494 (49.7%)
3	11 (3.3%)	164 (49.0%)	51 (15.7%)	226 (22.8%)
4	3 (.9%)	0 (.0%)	1 (.3%)	4 (.4%)
5	8 (2.4%)	2 (.6%)	4 (1.2%)	14 (1.4%)
6	3 (.9%)	0 (.0%)	3 (.9%)	6 (.6%)
7	3 (.9%)	1 (.3%)	0 (.0%)	4 (.4%)
Total	333	335	325	993

It is important to keep in mind that for the given constituents, the speakers may have been influenced by the precursor question. While the speakers were free to vary both NP syllable count and DP specificity in their responses, they heard a precursor question which always contained a mentioning of one of the referents: for subject focus, the object was mentioned, and for VP and object focus, the subject was mentioned. Therefore, it could be argued that the observations for the realizations of the given referents do not actually reflect an original speaker choice, but are merely repetitions of the DP and NP structure that was given to the speakers in the precursor question. To determine whether this is true, all realizations of given referents were compared with the realizations of these referents in the respective precursor questions. Table 7.5 gives the amount of overlap.

Table 7.5: Percentage of NP length and DP complexity outputs that mirror those of the respective precursor question.

	Subject NP length	Subject DP complexity	Object NP length	Object DP complexity
Precursor question	82.5%	97.5%	87.4%	55.4%

As can be seen from Table 7.5, the values are fairly high for the subject complexity. It appears that, when prompted for a sentence with object focus, the speakers most often repeated the structure of the subject constituent as it was presented to them in the precursor question. However, the picture for the object constituents is less clear cut. While the value for the object length yields a fairly high correspondence between precursor question and answer, the speakers only repeated the DP structure for given objects a little more than half of the time.

Therefore, it seems that the prompt from the precursor question cannot be the only factor to explain the choices made by the speakers in their realization of the referents. Rather, it appears that speakers systematically vary the syllable count of NPs to denote referents that are given or in focus. Also, it appears to be true at least for the object constituents that they use specific DP structures to refer back to a given referent, and indefinite constructions to introduce new referents into the context, which aligns well with cross-linguistic observations. Therefore, even if the structures in the precursor question align with those that are most prevalent in the answers by the speakers, this could simply be a reflection of the fact that a certain structure is more natural in these contexts.

7.3 Experiment 2: Phonetic marking of referents

The variation in the answer sentences that were recorded in the picture description paradigm meant that this data could not be analyzed acoustically to gain insight into the phonetic ways of marking information focus in Wenzhou. For this reason, a second experiment was designed, which used similar stimulus sentences as the first experiment, but controlled the speaker output structures more tightly. In accordance with the most commonly used method to investigate focus marking across Chinese dialects, the speakers were presented with written dialogues and asked to read them aloud in Wenzhou dialect.

7.3.1 Stimuli

The stimuli for the second experiment were question-answer pairs with target sentences that were similar in type to those of the first experiment. At the same time, the variation between the target sentences was reduced, so that all target sentences were SVO-sentences with one subject and one object in the active voice. The subject referent was always animate, and varied along three different lengths and three different tones (level, rising, falling), as can be seen in (6).

(6) **Subject referents**

	Hanzi	Citation forms	Tone	Translation
a.	阿妈	*a-**ma***	High level (33)	'mother'
b.	阿爸	*a-**pa***	High level (33)	'father'
c.	一个男	*i̍ kai **nɵ***	Low falling (31)	'a man'
d.	一个女	*i̍ kai **nø***	Low rising (24)	'a woman'
e.	一个男妹妹	*i̍ kai **nɵ** mai mai*	Low falling (31)	'a boy'
f.	一个女妹妹	*i̍ kai **nø** mai mai*	Low rising (24)	'a girl'

The object referent was varied between inanimate, animate, and human, and also carried either a level, falling, or rising tone, as shown in (7).

(7) **Object referents**

	Hanzi	Citation forms	Tone	Translation
a.	阿妈	*a-**ma***	High level (33)	'mother'
b.	阿爸	*a-**pa***	High level (33)	'father'
c.	(做)饭	*tsou **va***	Low level (11)	'(prepare) rice'
d.	(买)东西	*ma **mɵ-zz̩***	Falling (42.31)	'(buy) groceries'
e.	一个男	*i̍ kai **nɵ***	Low falling (31)	'a man'
f.	一个女	*i̍ kai **nø***	Low rising (24)	'a woman'
g.	一个牛	*i̍ kai **ŋau***	Low falling (31)	'an ox'
h.	一张桌(子)	*i̍ tɕi **dzu***	Low falling (31)	'a table'
i.	一杯茶	*i̍ bai **dzu***	Low falling (31)	'a (cup of) tea'
j.	一个男妹妹	*i̍ kai **nɵ** mai mai*	Low falling (31)	'a boy'
k.	一个女妹妹	*i̍ kai **nø** mai mai*	Low rising (24)	'a girl'

The subject and object references were used to compose 12 target sentences which were paired with four different context questions each. The context questions induced focus in different locations of the target sentence: on the subject (VP given), on the VP (subject given), on the object (subject given), or on the whole sentence (all new).[30] An example of a target sentence with the four precursor questions can be found in (8).

(8) **Target sentence:**

a.	一个女	正在	喝	一杯茶.		
	ʔi kai nø	*zẓ ta*	*ha*	*ʔi bai dzu*		
	NUM CL female	ASP	drink	NUM CL tea		
	‘A woman is drinking a cup of tea.’					
	Precursor questions:					
b.	Broad focus	你	说	什么	啊？	
		ni	*kuɔ*	*a ni*	*a*	
		you	say	what	Q	
		‘What did you say?’				
c.	Subject focus	什么	人	啊	正在	喝	一杯茶？[31]
		a ni	*naŋ*	*a*	*zẓ ta*	*ha*	*ʔi bai dzu*
		which	person	Q	ASP	drink	NUM CL tea
		‘Who is drinking a cup of tea?’					

[30] In order to keep both the question and the answer sentences as comparable as possible across the different focus conditions, the sentences in experiment 2 do not reflect the results of experiment 1, e.g. in terms of the definiteness of the referents in the questions and answers. The stimulus sentences of experiment 2 were still considered to be well-formed dialogues by the speakers who were recorded for this experiment.

[31] In place of the Pǔtōnghuà character 谁 shéi ‘who’, it is idiomatic to use the expression 什么人啊 in Wenzhou, which can be translated as ‘which person’. In grammars of Wenzhou, this expression is sometimes transcribed with different characters, e.g. 何样侬啊 in Hou 1998. Not all of the young speakers are familiar with these characters, but when presented with 什么人啊, they all produced the intended structure, which is why it has been used here. Similarly, the expression 什么啊 ‘what’ is transcribed as 何样啊 in Hou 1998, but the more common characters have been used here. The aspect particles 正在 are transcribed with the characters 著耷 in Hou 1998.

d.	VP focus	一个女	正在	作	什么	啊？
		ʔ̃ kai nø	*zẓ ta*	*tsu*	*a ni*	*a*
		NUM CL female	ASP	do	what	Q
		‘What is the woman doing?’				
e.	Object focus	一个女	正在	喝	什么	啊？
		ʔ̃ kai nø	*zẓ ta*	*ha*	*a ni*	*a*
		NUM CL female	ASP	drink	what	Q
		‘What is the woman drinking?’				

The target sentences were paired with four precursor questions each to create 48 question-answer pairs. These question-answer pairs were randomized, and alternated in a list with 66 other question-answer pairs from an unrelated experiment, to minimize the risk of the speakers interpreting a referent as given that occurred in another question-answer pair in the same experiment. A full list of question-answer pairs can be found in appendix 7.3.

7.3.2 Subjects

The subjects of the second experiment were 19 speakers (13 female) of the same age group as the speakers in the first experiment (mean age = 23;2, age range = 20-29). They were mostly high school graduates of the same high school in central Wenzhou as the speakers of the first experiment, and all born and raised in the central district of Lucheng Wenzhou. None of them reported to have lived outside of Wenzhou for a significant amount of time within the last 5 years, and all of them considered themselves fluent speakers of the Wenzhou dialect. Of the 19 speakers, eight recorded the stimuli sentences once, and eleven recorded all sentences twice. For those speakers with two recordings, the values were averaged over the recordings before statistical analysis.

7.3.3 Experimental procedure

Speakers were recorded in a sound-proofed recording studio in Wenzhou in individual sessions, and received a small payment for their participation. Each speaker was seated in front of a 13" monitor and given a Sennheiser pc130 headset. The experimenter ensured that the microphone of the headset was placed approximately 3 cm from the corner of the mouth of the subject. Via an external digitizer (UA-G1), the sound was recorded directly (44.1 KHz, 16 bits) on the laptop (Acer Aspire 1810TZ) on which the stimuli were displayed to the subject.

The speakers were first informed about the recording procedure. They were instructed to read out phrases and sentences presented on the screen using Wenzhou dialect, in a natural and clear fashion. If they were unsure how to pronounce a word or phrase, they could skip to the next item, and if they felt they had made a mistake, they could go back and repeat the recording of the previous item. They were told that they could interrupt or abort the recording at any point.

The recording itself was done using a script in the computer program PRAAT (Boersma & Weenink 2001).[32] This script would present the stimulus sentences in one by one, and record each stimulus individually after the speaker initiated the recording. Before the recording, all speakers completed a practice series with eight short phrases that were not part of the actual experiment. This was done in order to familiarize the speakers with the self-managed recording procedure. After completing the practice items, the speakers were asked to indicate whether they understood the recording procedure and were ready to start the experiment. They received a small payment for their participation.

7.3.4 Data analysis

After the recording, all utterances were checked for mistakes and long pauses or hesitations, and any such sentences were excluded from further analysis. All the remaining sentences were manually segmented, using acoustic and visual inspection of the sound wave and spectrogram. More specifically, the beginning and end of the subject and object constituents in all answer sentences were marked, and a PRAAT script extracted their duration, F_0 at 21 equidistant points within the constituent, and the position and F_0 value of the F_0 maxima and F_0 minima within the constituents.[33]

7.3.5 Results

7.3.5.1 Graphical analysis of results

As described in section 7.3.1, the experiment included lexical material with different tones on both the subject and the object constituents. For this reason, before a statistical analysis was conducted, it was checked whether the different tones were affected in a similar way by the different focus conditions. Below, a

[32] The script used for presenting and recording the stimuli was written by Jos Pacilly, and slightly modified by the author.

[33] The script used for segmenting and measuring the files was written by Jos Pacilly.

graphical analysis of the tonal contours on the subject and object referents is presented, as obtained by the measurement of F_0 on 21 equidistant points within the constituents. For the subject referents, five different lexical tones were recorded in the experiment (tone-bearing syllable marked in bold).

(9) **Target tones on subject referents**

	Tone type	Hanzi	Citation forms	Translation
a.	Level	阿**爸**	*a-**pa***	'father'
		阿**妈**	*a-**ma***	'mother'
b.	Short rising	一个**女**	*ʔi kai **nø***	'a woman'
c.	Long rising	一个**女**妹妹	*ʔi kai **nø** mai mai*	'a girl'
d.	Short falling	一个**男**	*ʔi kai **nɵ***	'a man'
e.	Long falling	一个**男**妹妹	*ʔi kai **nɵ** mai mai*	'a boy'

As Figure 7.6 shows, all five tested lexical tones on the subject constituent show a similar effect of the four different focus conditions. Regardless of the exact direction and alignment of the tonal contour, it always reaches a higher F_0 maximum under subject focus than under the other focus conditions, and covers a wider portion of the speakers' F_0 ranges. Furthermore, in all lexical tones it can be seen that the broad focus condition causes a wider F_0 range than the VP-focus and object focus conditions, and for the contour tones, it is also true that they reach a higher F_0 maximum in these conditions. Lastly, it seems to be true for all lexical tones that the VP focus and the object focus condition result in similar F_0 contours and excursion.

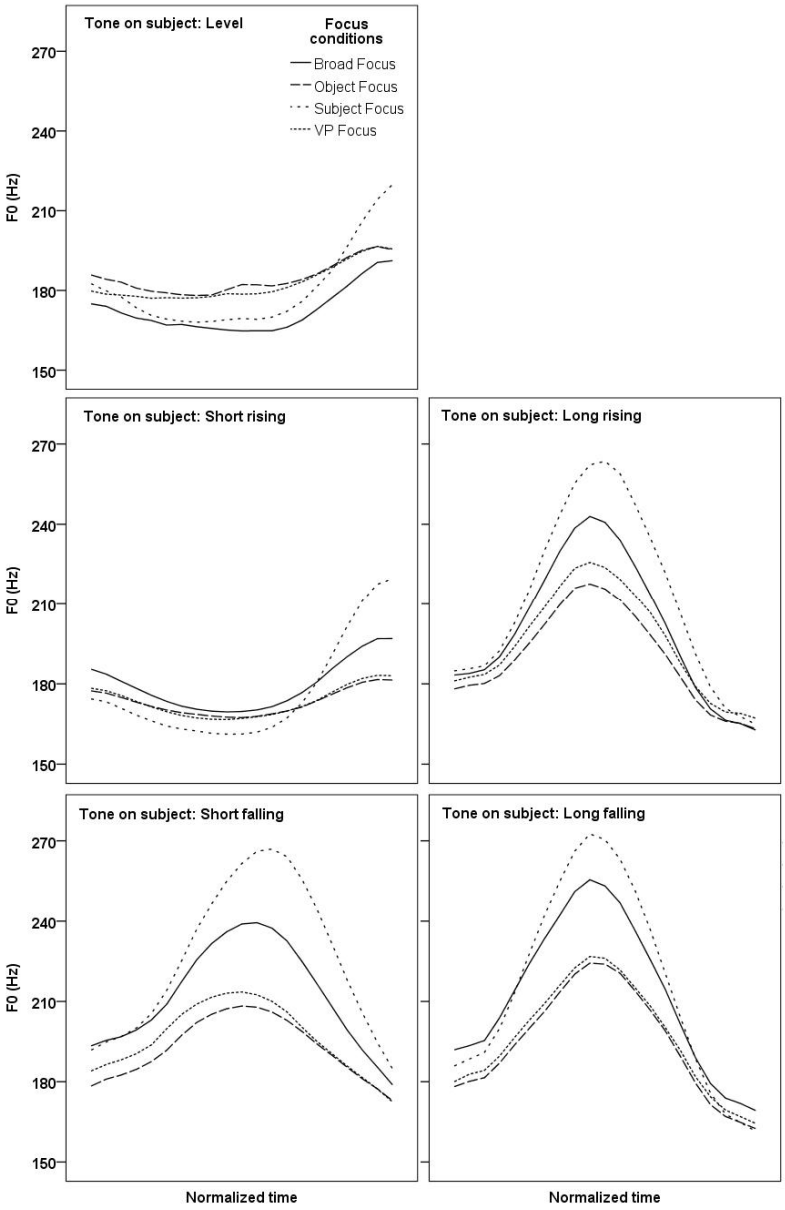


Figure 7.6: Averaged time-normalized tonal contours for subject referents.

Therefore, based on a graphical inspection of the F_0 contours for the subject referents, the following hypotheses can be stated, which will be statistically tested in the next section.

(10) **Hypotheses for subject referent realization**
a. F_0 range: subject focus > broad focus > VP focus, object focus
b. F_0 maxima: subject focus > broad focus > VP focus, object focus
c. F_0 minima: subject focus = broad focus = VP focus = object focus

For the object referent, the following tonal categories will be analyzed to make the analysis results comparable to that of the subject referent.

(11) **Target tones on object referents**

	Tone type	Hanzi	Citation forms	Translation
a.	Level	阿爸	*a-**pa***	'father'
		阿妈	*a-**ma***	'mother'
b.	Short rising	一个女	*i kai **nø***	'a woman'
c.	Long rising	一个女妹妹	*i kai **nø** mai mai*	'a girl'
d.	Short falling	一个牛	*i kai **ŋau***	'a cow'
		一张桌子	*i dʐoŋ **tɕu** zʐ*	'a table'
		一杯茶	*i bai **dzu***	'a cup of tea'
e.	Long falling	一个男妹妹	*i kai **nø** mai mai*	'a boy'

As can be seen from Figure 7.7, the influence of the different focus conditions on the tonal contour in the object constituents are not as clear-cut as for the subject constituent. There is a tendency, most clearly visible in the long falling and rising object tones, that the tones are scaled with a lower F_0 maximum and smaller F_0 excursion under subject focus (i.e. when the object was given) than under the three other focus conditions (when the object was new or in narrow focus).

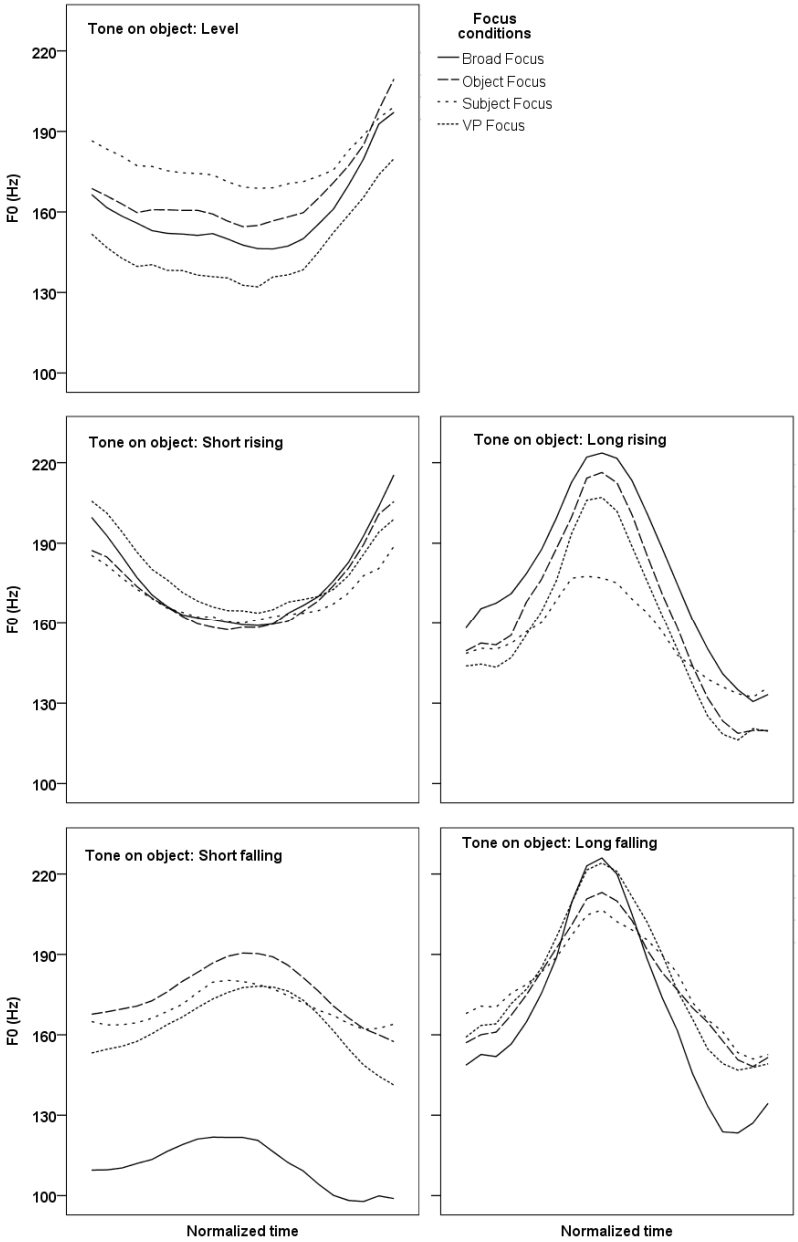


Figure 7.7: Averaged time-normalized tonal contours for object referents.

7.3.5.2 Statistical analysis of results

As described in the previous sections, it has become obvious from the inspection of the graphical results that all F_0 contours in the subject referents and some of the F_0 contours in the object referents were realized differently under different focus conditions. Specifically, it seems that the wh-focus on a constituent influences its F_0 maximum and F_0 range, but not the F_0 minimum. These three parameters will therefore be tested statistically, to determine how stable the observed effect is. Furthermore, the duration of the target constituents will be compared, since lengthening has been identified as another stable effect of focus in other dialects of Chinese.

Four measurements were subjected to statistical analysis for either constituent: F_0 maximum, F_0 minimum within the constituent, F_0 range (F_0 maximum – F_0 minimum), and the duration of the constituent. These measurements were subjected to both by-subjects ($F1$) and a by-items ($F2$) Repeated Measures (RM) ANOVA with FOCUS as dependent variable. In order to compare the effects of the four different FOCUS conditions (broad focus, object focus, subject focus, VP focus), a post-hoc pairwise comparison with Sidak-adjustment for multiple comparisons was performed. All reported degrees of freedom have been Huyhn-Feldt corrected when the requirement of sphericity was not met.

The two statistical analyses confirmed a main effect of FOCUS on all measurements in the subject constituents:

Duration: $F1(2.52,18) = 49.64$, $p < 0.001$, $F2(2.49,11) = 53.37$, $p < 0.001$
F_0 maximum: $F1(1.17,18) = 40.83$, $p < 0.001$, $F2(1.46,11) = 157.47$, $p < 0.001$
F_0 minimum: $F1(2.29,18) = 10.2$, $p < 0.001$, $F2(1.8,11) = 12.21$, $p < 0.001$
F_0 range: $F1(1.18,18) = 42.69$, $p < 0.001$, $F2(1.54,11) = 213.02$, $p < 0.001$

The same was true for the object constituents:

Duration: $F1(1.93,18) = 23.09$, $p < 0.001$, $F2(1.816,11) = 8.67$, $p < 0.01$
F_0 maximum: $F1(1.27,18) = 56.01$, $p < 0.001$, $F2(1.35,11) = 33.31$, $p < 0.001$
F_0 minimum: $F1(2.03,18) = 15.24$, $p < 0.001$, $F2(1.32,11) = 9.3$, $p < 0.01$
F_0 range: $F1(1.34,18) = 59.09$, $p < 0.001$, $F2(1.2,11) = 20.33$, $p < 0.001$

Table 7.6 summarizes the results of the post-hoc comparisons for all measurements, split by constituents and analysis type.

Table 7.6: Results of the post-hoc comparisons for all measurements, broken down by constituents and analysis type. BF = broad focus, OF = object focus, SF = subject focus, VPF = VP-focus. Significance level = $p < 0.05$.

Measurement	**By-subjects analysis**	**By-items analysis**
Duration subject	SF > BF > OF, VPF	SF > BF > OF, VPF
F_0 maximum subject	SF > BF > OF, VPF	SF > BF > OF, VPF
F_0 minimum subject	BF > OF, SF, VPF	BF > SF, OF, VPF
F0 range subject	SF > BF > OF, VPF	SF > BF > OF, VPF
Duration object	BF, OF, VPF > SF	BF, OF, VPF > SF
F_0 maximum object	OF > VPF BF, OF, VPF > SF	 BF, OF, VPF > SF
F_0 minimum object	BF, OF, VPF > SF	BF, OF, VPF > SF
F_0 range object	OF > VPF BF, OF, VPF > SF	OF > VPF BF, OF, VPF > SF

As the post-hoc comparisons confirm, the subject constituent is significantly longer, has a higher F_0 maximum and a wider F_0 range under subject focus than under the other three focus conditions. This is true for both the by-subjects and the by-items analysis, which differ very little from each other. Furthermore, the subject is also longer and has a higher F_0 maximum and F_0 range under broad focus (i.e. when the subject is new) than under object focus or VP focus (i.e. when the subject is given in the precursor question). This confirms a tripartite split for the subject in correspondence with three possible focus states: narrow focus > broad focus > given.

As for the F_0 minimum on the subject constituent, it is higher in broad focus than in all three other focus conditions. This speaks for a vertical expansion of F_0 range under focus that goes in both directions: it raises the F_0 maximum, and it lowers the F_0 minimum compared to the broad focus condition. At the same time, the F_0 minimum on the subject is also lowered when a narrow focus occurs later in the sentence. This is an interesting finding, because it shows that, at least for this measurement, a focus can even affect the implementation of a tone outside of its immediate focus domain.

For the object constituent, it is true for all measurements that they are lower in the subject focus condition than in all the other conditions. This means that the duration of the object is longer when the object is in focus, and shorter when the object is outside of the focus domain. Likewise, the object is scaled

with a higher F_0 maximum and F_0 minimum under focus, so it can be presumed that the speakers implement the tonal contours with less extreme and lowered F_0 excursions in a post-focal environment (cf. Xu 1999 and Chen 2010 for Standard Chinese). Again, it appears that when the object is given in the precursor question, as under subject focus, this affects its scaling and duration in the answer sentence. The F0 maximum and minimum results for the subject and the object are graphically represented in Figure 7.8.

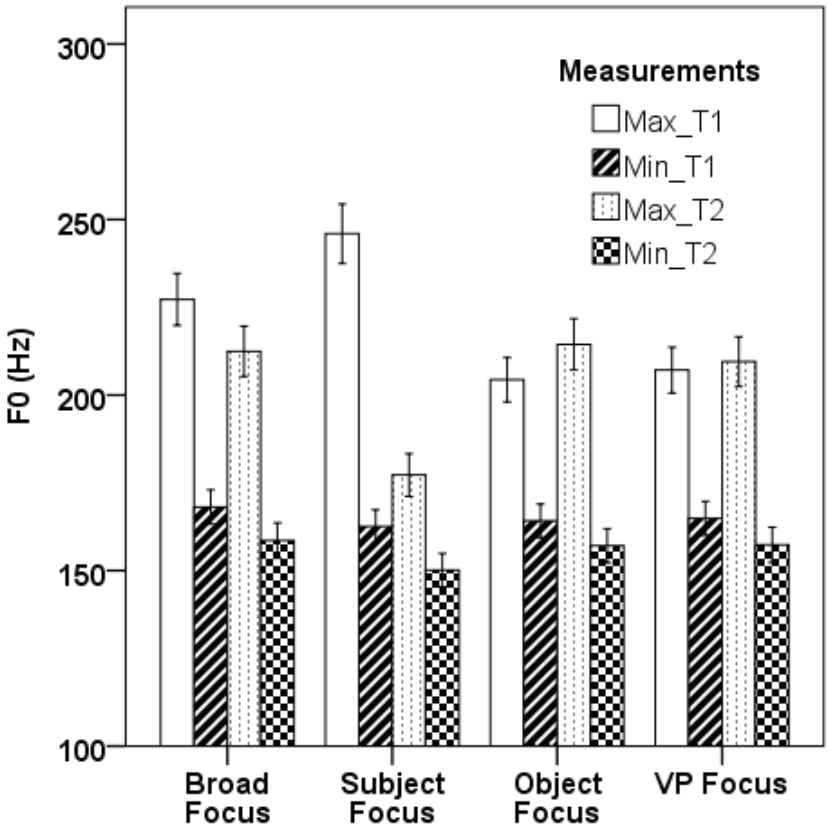


Figure 7.8: Averaged F_0 maximum and F_0 minimum values for the subject (T1) and object (T2) constituent, broken down by focus condition.

For the F_0 range measurements, the post-hoc comparisons furthermore show a significant difference in the implementation of (narrow) object focus in

comparison to VP focus: the F_0 range is significantly wider under object focus than under VP focus. The individual measurements, however, show that this effect is not very large: the F_0 min measurements are not significantly different between the two focus conditions, and the F_0 maximum measurements only reach significance in the by-subjects analysis. Nonetheless, it is interesting that there should be an effect of the width of the focus domain (the entire VP vs. just the object constituent), which translates to a difference in F_0 range. This observation aligns well with the earlier observation from the subject constituent, in which a ternary split between broad focus, narrow focus, and givenness can also be observed. The results are graphically represented in Figure 7.9.

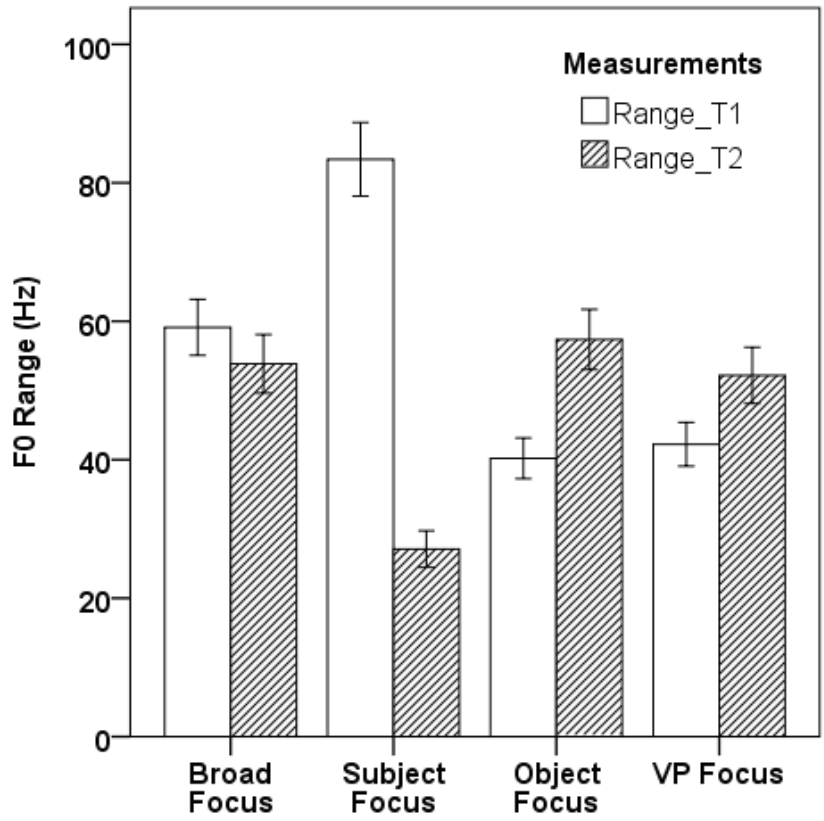


Figure 7.9: Averaged F_0 range values for the subject (T1) and object (T2) constituent, broken down by focus condition.

The duration measurements, as summarized in Figure 7.10, show the same tripartite split that has already been observed in the F_0 measurements. The subject constituent is significantly lengthened under subject focus, compared to all other focus conditions, and also significantly longer under broad focus than under VP focus and object focus. The object constituent on the other hand is of similar length in all three conditions where the object is within the focus domain (broad focus, VP focus, object focus), but significantly shorter under subject focus, where the object is given in the context question.

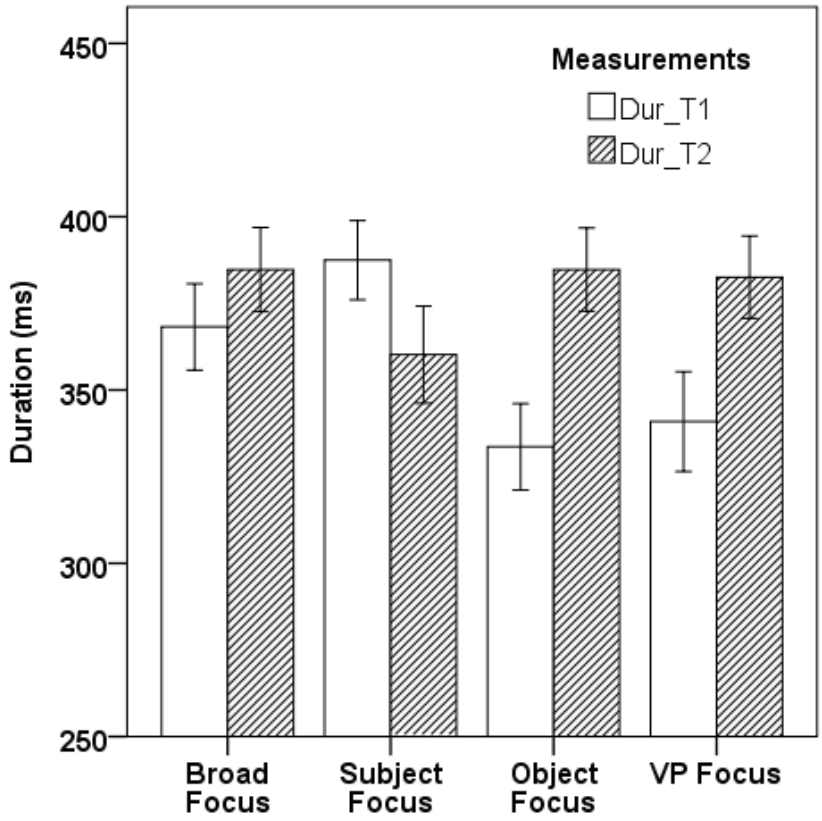


Figure 7.10: Averaged duration values for the subject (T1) and object (T2) constituents, broken down by focus condition.

7.4 Discussion

The contribution of the current study is twofold: it investigates how the speakers of Wenzhou Chinese mark different focus domains on the major constituents in sentences when they are in a more naturalistic discourse situation, and it also narrowly investigates the phonetic details of the realization of these focus domains when the speakers are more constrained in their expressions. In that sense, it illustrates some of the strategies that speakers have to mark the focus or givenness of constituents, and shows the preferences depending on the possibilities of the experimental setup.

In the first experiment described in this chapter, speakers were confronted with a discourse situation that mirrored a naturalistic dialogue rather closely. They heard a recording of a fellow Wenzhou speaker uttering a question, and they were asked to answer that question with the help of the information provided in a picture. As such, the experiment avoided the use of written language within the experiment completely, while still maintaining maximum comparability between the different utterances of the different speakers.

The analysis of the speaker strategies for the realization of the main sentential constituents (subject and object) shows that the speakers adjust their use of specific lexical options to the focus situation within the mini-discourse. In particular, the results show that they strategically and systematically add or omit syllables from the nominal heads, so that the syllable count increases when the constituent is under focus, and decreases when it is given in the precursor question and merely repeated in the answer. Even though the resulting syllable counts often overlap with those used in the precursor question, it is interesting to notice that speakers do deviate from what they heard in the question when they feel it is necessary to mark the referent for its information status.

This finding is well in line with previous findings on the realization of constituent focus in other dialects of Chinese, where it has been shown that speakers lengthen constituents under focus, compared to those that were out of focus or given. The current study shows that in situations where speakers have more freedom to choose the lexical material to construct their sentences, they deliberately use the manipulation of syllable count within the nominal head as a means to achieve the "lengthening" of constituents under focus.

Another strategy for focus/givenness marking investigated in the first experiment is the specific choice of DP structure that speakers make to realize focused or given referents in a discourse. As the results show, speakers have a convincing tendency to realize new or focused referents with the indefinite

numeral+classifier+noun structure, both in the subject and in the object constituents. For given referents, however, the strategies differ between subjects and objects: for given subjects, the speakers mostly choose to employ the definite classifier+noun structure, whereas for given objects, they predominantly use the more specific demonstrative+classifier+noun structure. However, all results are well in line with the expectations from theories on referent realization under focus. The speakers use a more definite structure to refer to familiar referents which have already been introduced into the discourse, and they prefer a more indefinite structure to denote referents that are new or focused.

In the case of the subject constituent, the results mirror to a large extent the DP structure that was used by the speaker who recorded the precursor questions, and there is a high overlap between the structure used in the questions and the structure used in the answers (97.5%). However, for the object constituent, the overall correspondence is much lower (55.4%), which shows that the speakers more often deviated from the DP structure in the question to mark the particular discourse status of the object constituent. Specifically, they predominantly used the demonstrative+classifier+noun structure to refer to object referents that were given in the precursor question, which shows that they understood the dialogue situation of the experiment well, and that they were aware that the referents in the answer had already been established in the prior discourse. This shows that speakers did not just parrot the structures from the precursor questions, but that they applied meaningful variation to the precise formulation of their answers, in accordance with the discourse needs in the specific situation.

For the second experiment, the speakers were more tightly constrained in their choices how to word the answers, so as to ensure a similar lexical and sentence structure across different discourse situations. This enabled a more fine-grained phonetic analysis of the realization of the two sentential constituents in question (subject and object) along different measurement parameters. For the second experiment, a further focus condition was added which was not tested in the first experiment, namely broad focus, in which the focus domain spanned the whole sentence.

The phonetic measurements revealed several clear tendencies for how the speakers manipulated the implementation of the tonal contours on the subjects and objects to convey the different focus contexts. In compliance with the results of the first experiment in this chapter and research on other dialects of Chinese, the speakers systematically and significantly lengthened the duration of constituents under focus, compared to those focus conditions where the

constituent was given in the precursor question. Furthermore, and this effect was more clearly visible on the subject than on the object, the speakers also lengthened the duration of constituents more when they were in narrow focus than when they were part of a wider focus domain (the sentence in case of the subject, and the VP in case of the object). This finding lends support to the assumption that speakers not only differentiate between the focus/non-focus status of constituents, but also take the given/not-given distinction into account.

The tripartite distinction between givenness, broad focus, and narrow focus was also visible in the F_0 measurements on both constituents. Considering the F_0 maximum and the F_0 range measurements, statistical analysis showed a tripartite distinction (narrow focus > broad focus > given) in all measurements on the subject constituent, and in all but one measurement on the object constituent (F_0 maximum was significant in the by-subject, but not in the by-item analysis). The F_0 minimum measurement was a bit less conclusive on the subject, where broad focus was singled out to have a higher F_0 minimum average than the other focus conditions (albeit by a small margin). For the object, however, also the F_0 minimum measurements aligned with the general trend and showed higher values under focus (broad focus, object focus, VP focus) than out of focus (subject focus).

The finding that all F_0 values on the object were uniformly lower and narrower in range under subject focus than under all three other focus conditions speak for a less distinct implementation of F_0 contours in a post-focal environment. At the same time, the distinction between sentence-wide focus (broad focus) and narrow focus (object focus/VP focus) does not reach significance on the object constituent. It seems that, for later constituents in the sentence, the effect of narrow focus is less extreme in extent (compare e.g. Xu 1999 for similar findings). Still, there is a (not statistically consistent) small effect of narrow object focus over the two wider focus options (broad focus, VP focus) to receive more distinct marking on the object, which again aligns with the tripartite distinction on the subject.

7.5 Conclusion

As reported in this chapter, two experiments were conducted to investigate the realization of subject and object referents under different focus conditions in Wenzhou Chinese. The first experiment was set up in a way to allow the speakers maximal freedom of expression, while still preserving the comparability of the realizations of the target sentences to the greatest amount

possible. To this end, speakers were presented with acoustic precursor questions, and they were free to phrase their answers with any lexical material they liked, within the limits of a picture description task that specified the intended content of the target sentences.

Analysis of the complexity and structure of the DPs used by the speakers to refer to the referents in the different focus conditions revealed that speakers tend to choose longer lexical material (i.e. more syllables) to realize referents under focus than to realize referents that are already given in the precursor question. This observation holds true for both subject and object referents. In terms of DP structure, the speakers tend to use a definite classifier+noun structure to refer to given subject constituents, mirroring largely the structure that was used in the precursor question to introduce these referents. For the focused referents, they predominantly chose an indefinite numeral+classifier+noun structure for both subjects and objects. Interestingly, for object referents that were given, they preferred an overall different structure, namely demonstrative+classifier+noun, which in many cases was different from the DP structure that the given object referents were introduced with in the precursor question. This speaks for a high sensitivity of the speakers for the different discourse situations, which influence their choice of wording for the individual constituents.

To investigate in detail the phonetic means that speakers of Wenzhou Chinese have at their disposal to mark the discourse status of referents, a second experiment was conducted, which controlled the lexical material in the target dialogues more closely. Analysis of the target (answer) sentences across four different discourse conditions (broad focus, subject focus, object focus, VP focus) revealed that speakers systematically make use of duration and F_0 parameters to mark the referents with respect to the parameters focused, new, and given.

For the subject constituents, the measurements showed a clear tripartite division between the realizations of this constituent, with subjects systematically being longer and having a higher and wider F_0 scaling under subject focus, compared to the other three focus conditions. Furthermore, they were also longer and realized with a higher and wider F_0 und broad focus (all-new) than under object and VP focus (subject given). This tripartite distinction between focused, new, and given closely mirrors the results of Féry & Kügler 2008 for German, where a similar tripartite distinction was found.

For the object constituent, the main division that was found back in all measurements was that between subject focus and all other focus conditions. It

was found that the object constituent was systematically shorter and scaled overall lower and with less F_0 excursion when following a focused subject, than when it was new or focused itself. For the objects, the distinction between "new" and "focused" only held for the F_0 range and F_0 maximum measurements, and only came out statistically significant when comparing object focus to VP focus. A possible explanation for this can be found in the overall reduced excursion size of the focus effect later in the sentence, which has also been observed for Standard Chinese before (Xu 1999). The significance of the F_0 range effect between object focus and VP focus suggests that there is a difference between "new" and "focused" also for the object, but that the overall smaller excursion size of the focus effect prevents it from systematically reaching significance.

Chapter 8

Conclusion

This thesis set out to experimentally investigate the relation between tonal realization, prosodic phrasing, and focus realization with data from the Wenzhou dialect of Chinese. Specifically, the following research questions were addressed, as listed in section 1.2.2 of this thesis:

Tone sandhi and prosody (Chapter 3):

- Which factors determine the application of tone sandhi in disyllabic targets which are ambiguous between two prosodic structures?

Tone sandhi and focus (Chapters 3 and 4):

- Can the presence of contrastive focus in (only one of the syllables of) a disyllabic lexical compound block lexical tone sandhi?
- If not, how are the acoustic reflexes of focus distributed within the disyllabic lexical compound if only one of the syllables is focused, compared to focus on the entire disyllabic lexical word?

Tonal realization and prosodic structure (Chapter 5):

- Is the implementation of tonal contours affected by prosodic structure? If yes, which component of prosodic structure (prosodic boundaries/prosodic heads) is more important for the way tonal contours are implemented?

Tonal realization, prosodic structure, and focus (Chapter 5):

- Is the effect of prosodic structure on tonal implementation identical to the effect of focus?

Tone sandhi contour implementation and prosodic structure (Chapter 6):

- How are tonal contours implemented/scaled in sentences with different numbers of words per constituent? Is the scaling of the contours based on sentence or on constituent length?
- How does syntactic embedding affect the scaling of tonal contours? Which level of syntactic complexity is reflected in the tonal scaling?

Tonal realization and focus/givenness (Chapter 7):

- Do the speakers of Wenzhou use lexical means to mark referents in different discourse situations?
- Is there a difference in tonal realization between constituents that are given, broadly focused, and narrowly focused?

A summary of the experimental findings that were presented in the individual chapters will be given in section 8.1. In section 8.2, these findings will be set into relation with the broader research questions outlined in the introduction. Section 8.3 presents suggestions for further research.

8.1 Summary of experimental results

Chapter 3 was concerned with the application conditions for tone sandhi in disyllabic structures that are ambiguous with respect to lexical wordhood. Two theoretical claims were tested, namely whether tone sandhi application in disyllabic verb-object constructions correlates with their degree of lexicalization, and that the presence of contrastive focus influences the application likelihood of tone sandhi. Additionally, it was tested whether contextual factors, such as the presence of a carrier sentence or the co-elicitation of disyllabic lexical compounds, would influence the application of tone sandhi in the verb-object constructions.

As was shown with experimentally obtained data, only the first claim could be confirmed for the speech of the young Wenzhou speakers. The degree of lexicalization of verb-object constructions, as measured according to semantic, syntactic, and morphological criteria, positively correlated with the relative magnitude of tone sandhi application in the realizations by the speakers. On the other hand, recording the disyllabic verb-object constructions in sentence-medial position and in contrastive focus context did not significantly influence the number of instances of tone sandhi application.

However, the experiment reported in Chapter 3 found that the most important predictor for whether the disyllabic verb-object constructions are realized as tone sandhi contours or with phrasal prosody was actually the prosodic context. More specifically, the experimental results showed that speakers were much more likely to realize verb-object constructions as phrases when they were presented in the context of other verb-object constructions. On the other hand, when the verb-object constructions were presented in the context of disyllabic compounds, the speakers most often applied tone sandhi to both the

disyllabic compounds and the disyllabic verb-object constructions. This speaks for a high variability in the tone sandhi application behavior of the young speakers, while it could be shown that the contextual factors that influence the application of tone sandhi are different from what has been assumed in previous literature.

In Chapter 4, the research question concerned the phonetic effects of contrastive focus on the tonal contours that result from lexical tone sandhi. Particularly, the experimental setup varied the position and extent of the focus domain with respect to the disyllabic tone sandhi domain, such that the focus domain would either precede, follow, undercut, or encompass the tone sandhi domain. In that way, it was tested whether (i) the presence of focus on a sub-part of the tone sandhi domain can interrupt the application of the tone sandhi process itself, and (ii) if not, whether the phonetic implementation of the tone sandhi contour would be different under focus on the entire tone sandhi domain, compared to focus only on one of its syllables.

Concerning (i), it was found, in agreement with the findings in Chapter 3, that the presence of focus did not affect the application of tone sandhi on the lexical compounds. Extending the findings of Chapter 3, it was shown that tone sandhi, which presumably serves as a marker for lexicalization, even applies in contexts where the speakers want to stress the importance of one of the syllables of the compound over that of the other syllable. Therefore, focus on a constituent below the word level cannot affect the application of phonological processes in Wenzhou (unlike e.g. pitch accent assignment on the stressed syllable in Dutch and English, which can be overridden by focus requirements).

Additionally, it was shown that even on the phonetic level, focusing only one of the syllables of the compound did not consistently lead to an F_0 or duration difference when compared to focus on the entire word. This means that, even on the phonetic level, the tone sandhi contour is only affected by focus as one whole, and its components are not individually accessible to focus marking. Instead, the focus effect (lengthening and F_0 range expansion) is distributed over the entire tone sandhi domain as a whole, and neither cue is sufficient to differentiate e.g. focus on the first syllable from focus on the second syllable, or from focus on the whole word. These findings underline the importance of the tone sandhi domain as phonological domain, and its special status compared to disyllabic domains in other dialects which do not have tone sandhi.

In Chapter 5, the influence of prosodic structure and focus on tonal realization was tested. Based on hypotheses from findings on the segmental level, which predict a strengthening of articulation in prosodically strong and focused

positions, it was investigated whether a similar strengthening effect could be observed for the implementation of tonal contours. Specifically, it was tested to what extent tonal contours are susceptible to the influence of neighboring tonal targets in the context of the two influence factors (prosodic structure and focus).

Moreover, in order to tease apart the influence of prosodic boundaries from the influence of prosodic headedness, two syntactic structures were compared which differ in both of these prosodic characteristics. In verb-object structures, both components form a prosodic phrase with each other, but due to the principle of nonhead prominence, only the object bears prosodic prominence. In adverb-verb structures on the other hand, both components constitute their own prosodic phrase and thereby both acquire prosodic headedness. Therefore, the two structures differ both with respect to the prosodic boundary between the components (V-O: Prosodic word boundary, ADV-V: Prosodic phrase boundary), and with respect to the prosodic prominence distribution (V-O: Prosodic head = Object, ADV-V: Prosodic head = Both).

Comparing the amount of coarticulatory influence between the two structures, it was found that the tonal trajectories of rising and falling tones were significantly steeper for the verbs in verb-object structures than for the adverbs in adverb-verb structures. On the other hand, no significant difference in the steepness of the contours was found between the objects in verb-object structures and the verbs in adverb-verb structures. These findings were interpreted to indicate that the relevant component of prosodic structure, which influences tonal coarticulation in Wenzhou Chinese, is prosodic headedness. Specifically, tones are articulated more autonomously and with steeper contours when they are in prosodically strong positions, but they are more susceptible to the influence of adjacent tonal targets when they are in prosodically weak positions.

By comparing this effect of prosodic prominence on tonal implementation to the effect of focus, it was observed that the two effects are not identical. Rather, while tonal contours in prosodically weak positions showed less influence of adjacent tonal targets under focus, tonal contours in prosodically strong positions showed further strengthening under focus. These findings are incompatible with theories in which focus is implemented directly as prosodic prominence, and in which the only requirement of focus for the grammar is to have a prosodically strong position on the location of the focused constituent. If focus were implemented as prosodic prominence, it would be unclear how the observed further strengthening of prosodically strong positions would be conceptualized.

Instead, the findings lend further support to theories which account for the effect of focus as a strengthening of tonal implementation that is independent of prosodic structure. In such a theory, prosodically strong positions and focus both cause a difference in the implementation of tonal contours, but do so independently of each other. This would predict that the two effects may be (partially) cumulative, which is in line with the experimental findings for Wenzhou Chinese as presented in Chapter 5.

Chapter 6 looked at the properties of tonal realization on the sentence level. In order to investigate whether and how prosodic structure influences the implementation of tonal contours in a more global manner, sentences consisting of rise-fall tone sandhi contours were investigated. By keeping the tonal properties of the individual words constant, it was attempted to make the Wenzhou test sentences comparable to earlier investigations in African tone languages and intonational languages, where F_0 scaling was investigated on the basis of the peak scaling of pitch accent or tonal peaks.

In contrast to these languages investigated earlier, where sentential F_0 scaling was found either to be pre-planned globally based on sentence length, or implemented locally from one constituent to the other, it was found for Wenzhou that F_0 scaling was prosodically mediated. Specifically, it was found that a manipulation of the length of the subject or object constituent affected the scaling of the initial peak within that constituent, but not within the respective other constituent. This indicates that F_0 pre-planning in Wenzhou is performed on a semi-global level, namely that of the syntactic/prosodic phrase. At the same time, the location of the F_0 reset was found to have a fixed location in the structure, which indicated that the speakers do not re-adjust the prosodic structure of a sentence to balance the length of the individual constituents, as has been found for some Romance languages.

Testing the F_0 scaling of embedded clauses, a second finding in Chapter 6 was that the scaling is sensitive to the syntactic complexity of the embedded structure. While it is true for all test sentences that the embedded clause was also prosodically embedded (i.e. scaled lower than the matrix clause), the F_0 difference between matrix verb and embedded clause was larger for an embedded CP than for an embedded VP. This difference was found regardless of whether the embedded subject was overtly spelled out or not. In that sense, it can be concluded that F_0 scaling in Wenzhou Chinese is used as a marker for important syntactic differences to distinguish structurally different sentences with similar linear word order.

Finally, Chapter 7 investigated the marking of wh-focus and givenness, both in the lexical and in the phonetic/phonological respect. For the test of lexical complexity, the answers from the speakers were elicited with the help of a picture description task, which allowed the speakers considerable freedom in their realization of the focused and given referents. It was found that, similar to other languages, speakers use shorter lexical, and more definite, forms to talk about given referents in a discourse, and longer lexical and indefinite forms to talk about focused, not previously introduced referents.

In a second experiment, which investigated the phonetic marking of focus and givenness in a more experimentally controlled way, it was shown that speakers systematically mark given referents differently from focused referents, both in terms of F_0 and in terms of duration. In addition, the experiment also found a difference in F_0 and duration marking between referents in broad focus and those in narrow focus context. These findings corroborate accounts of information structure which stress that the complexity of referent marking cannot just be accounted for in terms of presence/absence of focus. Rather, the givenness of a referent can act as an additional factor, and induce a tripartite division in the realization of the respective tonal contours on the referents.

8.2 General conclusions

From the experimental results presented in the previous section, several important conclusions can be drawn.

First of all, the tone sandhi application observed the young speakers proved to be more variable than assumed in the previous literature in some respects, and at the same time more stable in other respects. A disyllabic collocation of two monosyllabic lexical words can be treated as one word, and consequently be realized with the lexical tone sandhi contour, or be treated as two separate words, and be realized in a way that is consistent with phrasal prosodic requirements. The exact realization of such an ambiguous structure that is chosen by a speaker in a certain moment can be predicted to some extent based on the lexical properties of the collocation. However, the largest influence factor seems to be whether the disyllabic collocation is uttered in the context of other (clearly lexicalized) compounds, or together with other phrasal structures.

In that sense, tone sandhi serves as a lexicalization marker, but can also assume other functions in ambiguous structures, and be implemented analogously to surrounding tonal contours. This finding is difficult to reconcile with theories in which the tone sandhi application domain is crouched in the

framework of prosodic levels, which are derived from syntactic structure (e.g. Chen 2000). Clearly, for the young Wenzhou speakers, its syntactic composition is not the sole criterion for the application tone sandhi on a disyllabic structure. Rather, some structures can be ambiguous, and in these ambiguous structures, several factors which are not related to prosody play a role in determining the precise application rate of tone sandhi.

Once the tone sandhi contours are derived on the word level, however, their implementation on the phrase and sentence level can be influenced both by prosody and by focus. For the former, the scaling of the tonal contour, i.e. its relative height within the speaker's pitch range, is dependent on the size of the phrasal prosodic constituent in which the particular word appears, and on its position within that constituent. Similarly, the height of an F_0 peak, compared to the preceding F_0 peak, is related to properties of the underlying syntactic structure, and gives cues to the listener about e.g. the complexity of structure in an embedded clause.

As for focus, the realization of the tone sandhi contour is magnified, similarly to the F_0 expansion effect of focus on tones on lexical monosyllables. However, the focus effect is distributed more or less evenly over the entire disyllabic contour, even when only one of the syllables within the disyllabic tone sandhi domain is the precise location of focus. It appears that under focus, the requirement to mark the exact focus location is in conflict with the requirement to treat the entire disyllabic lexical domain as one whole for the sake of tonal realization. The speakers resolve that conflict by giving precedence to the preservation of the disyllabic coherence of the tonal contour, at the expense of precisely marking the exact location of focus.

Therefore, the tone sandhi contour cannot be broken apart by focus, neither in the phonological target selection (i.e. by blocking tone sandhi), nor in the phonetic implementation of the selected target (i.e. by locating the focus effect on the focused syllable alone). The effect of focus, which acoustically manifests itself in lengthening and F_0 range expansion, applies on the tone sandhi domain in the same way that it would apply on lexical tones, and it treats the disyllabic tone sandhi contour as a single tonal contour.

On monosyllabic words, it can be seen that F_0 range expansion is not the only acoustic reflex of focus. On contour tones, focus also affects the strength of the tonal realization, so that the tones are less influenced by adjacent tonal targets. This effect is similar in kind, but independent of the prosodic strengthening effect, which affects monosyllabic tonal targets in prosodic head positions. This thesis has shown that the two strengthening effects of focus and

of prosody are cumulative, and that therefore one cannot be explained away with the help of the other.

Another piece of evidence, which speaks for an independence of the focus effect from prosodic phrasing, is its gradience. Speakers seem to be able to adjust the magnitude of focal strengthening, for example to distinguish broadly focused from narrowly focused constituents, and at the same time differentiate both types from given constituents. Outside of lexical means of referent denotation, the phonetic implementation of tonal targets therefore represents another powerful tool for the speakers to convey detailed communicative distinctions to listeners.

8.3 Directions for future research

On the basis of the findings laid out in this thesis, several other aspects of tone realization, prosodic phrasing, and focus can be investigated.

As for prosodic phrasing, it could be tested whether the observed strengthening effect of prosodic headedness also holds for tone sandhi tones. It could be argued that the flattening of lexical tonal contours in prosodically weak positions, as observed in this thesis, could stand in potential conflict with the characteristics of the tone sandhi contours. For example, it has been suggested for Taiwanese that tonal coarticulation is minimized in order to maintain the distinguishability of the tone sandhi tonal contours (Lin 1988). For Wenzhou, it would be interesting to test how speakers would realize tone sandhi contours, for which contour recognizability is crucial, in prosodically weak positions, in which the distinct realization of contours might be compromised.

As for F_0 scaling, several further aspects could be explored. For example, in addition to the subordinated clauses tested in this thesis, it could be investigated whether and how the syntactic structure of coordinated clauses is reflected similarly in the intonational implication. Additionally, it could be tested how the observed F_0 scaling properties in the Wenzhou sentences are affected by the presence of focus. Particularly, it would be interesting to observe whether focus on a certain constituent would induce an effect on the F_0 scaling that is different from the prosodic effect of a boundary. If this were the case, the argumentation that prosodic structure and focus effects are in principle independent would be further corroborated.

As for focus, it is an open research question whether wh-induced focus and contrastive focus have the same phonetic reflexes. In light of the acoustic complexity of the focus effects that were found for Wenzhou in this thesis (e.g.

the acoustic distinction between givenness, broad focus, and narrow focus of constituents as presented in Chapter 7), it could be expected that Wenzhou might represent a good test case to corroborate findings for languages like English (Katz & Selkirk 2011). One research area that has been completely neglected in this thesis is the information-structural notion of *topic*, and any focus effects, whether brought about by wh-focus or contrastive focus, could also be compared to the effects of topic, as in Chen 2009; Wang & Xu 2006, 2011.

Finally, while much research has already been devoted to the analysis of tonal realization in other dialects of Chinese, it can sometimes be difficult to directly compare the observed effects across different dialects, because each study uses its own methodology and stimulus composition. Investigations which directly compare different dialects (as in e.g. Chen et al. 2009; Xu et al. 2012), especially if they are expected to have very dissimilar tonal properties, can provide important insights in the abstract mechanisms that underlie the effects in question.

References

Alexiadou, Artemis. 1997. *Adverb Placement: A Case Study in Antisymmetric Syntax*. Amsterdam: John Benjamins.

Arvaniti, Amalia. 2003. Peak scaling in Greek and the role of declination. *Proceedings of the 15th International Congress of Phonetic Sciences*, Barcelona. 2269-2272.

Avesani, Cinzia. 1987. Declination and sentence intonation in Italian. *Quaderni del Laboratorio di Linguistica della Scuola Normale Superiore di Pisa* 1(1). 8-23.

Baltazani, Mary. 2006. Focus, prosodic phrasing, and hiatus resolution in Greek. *Laboratory Phonology 8*, ed. by Louis Goldstein, D.H. Whalen and Catherine T. Best, 473-494. Berlin/New York: Mouton de Gruyter.

Baumann, Stefan, Johannes Becker, Martine Grice, and Doris Mücke. 2007. Tonal and articulatory marking of focus in German. *Proceedings of the 16th International Congress of Phonetic Sciences*, Saarbrücken. 1029-1032.

Baumann, Stefan, Martine Grice, and Susanne Steindamm. 2006. Prosodic Marking of Focus Domains - Categorical or Gradient? *Proceedings of the 3rd International Conference on Speech Prosody*, Dresden. 301-304.

Bishop, Jason. 2010. Information Structural Expectations in the Perception of Prosodic Prominence. *UCLA Working Papers in Phonetics* 108(1). 203-225.

Boersma, Paul, and David Weenink. 2001. Praat, a system for doing phonetics by computer. *Glot International* 5(9/10). 341-345.

Breen, Mara, Evelina Fedorenko, Michael Wagner, and Edward Gibson. 2010. Acoustic correlates of information structure. *Language and Cognitive Processes* 25(7). 1044-1098.

Breen, Mara, Duane G. Watson, and Edward Gibson. 2011. Intonational phrasing is constrained by meaning, not balance. *Language and Cognitive Processes* 26(10). 1532-1562.

Brooke, Julian, Greg Coppola, T.-Y. Vivian Lee, and Yang Zhao. 2009. Phrasing, Tempo, and Focus: A Prosodic-Hierarchy Approach to Mandarin Tone Sandhi Domains. *Simon Fraser University Working Papers in Linguistics* 2 (Proceedings of the 2nd International Conference on East Asian Linguistics)(1).

Bruce, Gosta. 1977. *Swedish word accents in sentence perspective*. Lund: Geerlup.

Büring, Daniel. 2006. Been there, marked that - A tentative theory of second occurrence focus. Los Angeles: University of California Los Angeles, ms.

Büring, Daniel. 2010. Towards a Typology of Focus Realization. *Information Structure. Theoretical, Typological, and Experimental Perspectives*, ed. by Malte Zimmermann and Caroline Féry, 177-205. Oxford: Oxford University Press.

Byrd, Dani, and Elliot Saltzman. 1998. Intragestural dynamics of multiple prosodic boundaries. *Journal of Phonetics* 26(2). 173-199.

Cabré, Teresa, and Pilar Prieto. 2005. Positional and metrical prominence effects on vowel sandhi in Catalan. *Prosodies. With Special Reference to Iberian Languages*, ed. by Sónia Frota, Marina Vigário and Maria João Freitas, 123-158. Berlin/New York: Mouton de Gruyter.

Cao, Jianfen. 1982. 常阴沙话古全浊声母的发音特点—吴语清浊音辨析之一 [The Characteristics of the Ancient Initial Voiced Consonants in Changyinsha Dialect]. *中国语文 [Chinese language]* 167(4). 273-278.

Cao, Jianfen, and Ian Maddieson. 1992. An exploration of phonation types in Wu dialects of Chinese. *Journal of Phonetics* 20(1). 77-92.

Cao, Jianfen, and Yuling Zheng. 2006. Articulatory strengthening and prosodic hierarchy. *Proceedings of the 3rd International Conference on Speech Prosody*, Dresden.

Chang, Yueh-Chin, and Feng-Fan Hsieh. 2012. Tonal coarticulation in Malaysian Hokkien: A typological anomaly? *The Linguistic Review* 29(1). 37-73.

Chao, Yuen Ren. 1928. *Studies in the Modern Wu Dialects.* Beijing: Qinghua University.

Chao, Yuen Ren. 1948. *Mandarin primer*. Cambridge, MA: Harvard University Press.

Chao, Yuen Ren. 1967. Contrastive Aspects of the Wu Dialects. *Language* 43(1). 92-101.

Chao, Yuen Ren. 1968. *A Grammar of Spoken Chinese*. Berkeley: University of California Press.

Chen, Matthew Y. 1987. The syntax of Xiamen tone sandhi. *Phonology Yearbook* 4. 109-149.

Chen, Matthew Y. 1991. An Overview of Tone Sandhi Phenomena Across Chinese Dialects. *Languages and Dialects of China*, ed. by William S-Y. Wang, 113-158. Berkeley: Project on Linguistic Analysis.

Chen, Matthew Y. 2000. *Tone Sandhi. Patterns across Chinese dialects*. Cambridge Studies in Linguistics 92. Cambridge: Cambridge University Press.

Chen, Szu-wei, Bei Wang, and Yi Xu. 2009. Closely related languages, different ways of realizing focus. *Proceedings of the 10th annual conference of the International Speech Communication Association (Interspeech)*, Brighton. 1007-1010.

Chen, Yiya. 2004. Focus and Intonational Phrase Boundary in Standard Chinese. *Proceedings of the 4th International Symposium on Chinese Spoken Language Processing (ISCSLP)*, Hong Kong. 41-44.

Chen, Yiya. 2006. Durational adjustment under corrective focus in Standard Chinese. *Journal of Phonetics* 34(2). 176-201.

Chen, Yiya. 2008. The acoustic realization of vowels of Shanghai Chinese. *Journal of Phonetics* 36(4). 629-648.

Chen, Yiya. 2009. Prosody and information structure mapping: Evidence from Shanghai Chinese. *Chinese Journal of Phonetics* 2(1). 123-133.

Chen, Yiya. 2010. Post-focus F0 compression - Now you see it, now you don't. *Journal of Phonetics* 38(4). 517-525.

Chen, Yiya. 2011. How does phonology guide phonetics in segment - f0 interaction? *Journal of Phonetics* 39(4). 612-625.

Chen, Yiya, and Carlos Gussenhoven. 2008. Emphasis and tonal implementation in Standard Chinese. *Journal of Phonetics* 36(4). 724-746.

Chen, Yiya, and Yi Xu. 2006. Production of Weak Elements in Speech - Evidence from F0 Patterns of Neutral Tone in Standard Chinese. *Phonetica* 63(1). 47-75.

Chen, Yiya, and Jiahong Yuan. 2007. A Corpus Study of the 3rd Tone Sandhi in Standard Chinese. *Proceedings of the 8th annual conference of the International Speech Communication Association (Interspeech)*, Antwerp. 2749-2752.

Cheng, Lisa Lai Shen. 1987. On the Prosodic Hierarchy and Tone Sandhi in Mandarin. *Toronto Working Papers in Linguistics* 7(1). 24-52.

Cheng, Lisa Lai Shen, and Rint Sybesma. 1999. Bare and Not-So-Bare Nouns and the Structure of NP. *Linguistic Inquiry* 30(4). 509-542.

Cheng, Lisa Lai Shen, and Rint Sybesma. 2005. Classifiers in four varieties of Chinese. *The Oxford Handbook of Comparative Syntax*, ed. by Guglielmo Cinque and Richard S. Kayne, 259-292. Oxford: Oxford University Press.

Chi, Teelee R. 1985. *A Lexical Analysis of Verb-noun Compounds in Mandarin Chinese*. Taipei: Literary Crane Publishers.

Cho, Taehong. 2004. Prosodically conditioned strengthening and vowel-to-vowel coarticulation in English. *Journal of Phonetics* 32(2). 141-176.

Cho, Taehong. 2006. Manifestation of Prosodic Structure in Articulation: Evidence from Lip Kinematics in English. *Laboratory Phonology VIII: Varieties of Phonological Competence*, ed. by Louis Goldstein, D.H. Whalen and Catherine T. Best, 519-548. Berlin/New York: Mouton de Gruyter.

Cho, Taehong. 2011. Laboratory Phonology. *Continuum Companion to Phonology*, ed. by Nancy C. Kula, Bert Botma and Kuniya Nasukawa, 343-368. London/New York: Continuum.

Cho, Taehong, and Patricia Keating. 2009. Effects of initial position versus prominence in English. *Journal of Phonetics* 37(4). 466-485.

Chow, Ivan. 2006. Interactions between Syllabic-Level Prosody and Prosodic-Group Boundary Markers in Cantonese. *Proceedings of the Annual Conference of the Canadian Linguistic Association*, York.

Cinque, Guglielmo. 1993. A Null Theory of Phrase and Compound Stress. *Linguistic Inquiry* 24(2). 239-297.

Cinque, Guglielmo. 2004. Issues in adverbial syntax. *Lingua* 114(6). 683-710.

Connell, Bruce. 2004. Tone, Utterance Length and F0 Scaling. *Proceedings of the International Symposium on Tonal Aspects of Languages: With Emphasis on Tone Languages*, Beijing. 41-43.

Costa, Joao, and Nancy C. Kula. 2008. Focus at the interface: Evidence from Romance and Bantu. *The Bantu-Romance Connection: A comparative investigation of verbal agreement, DPs, and information structure*, ed. by Cécile de Cat and Katherine Demuth, 293-322. Amsterdam: John Benjamins.

D'Imperio, Mariapaola, Gorka Elordieta, Sónia Frota, Pilar Prieto, and Marina Vigário. 2005. Intonational phrasing in Romance: The role of syntactic and prosodic structure. *Prosodies. With Special Reference to Iberian Languages*, ed. by Sónia Frota, Marina Vigário and Maria João Freitas, 59-89. Berlin/New York: Mouton de Gruyter.

Dai, John Xiang-Ling. 1998. Syntactic, phonological, and morphological words in Chinese. *New Approaches to Chinese Word Formation: Morphology, Phonology and the Lexicon in Modern and Ancient Chinese*, ed. by Jerome Packard, 103-134. Berlin/New York: Mouton de Gruyter.

de Jong, Kenneth, Mary E. Beckman, and Jan Edwards. 1993. The interplay between prosodic structure and coarticulation. *Language and Speech* 36(2-3). 197-212.

Dehé, Nicole, and Vieri Samek-Lodovici. 2009. On the prosody and syntax of DPs: evidence from Italian noun adjective sequences. *Natural Language and Linguistic Theory* 27(1). 45-75.

Dresher, B. Elan. 1994. The Prosodic Basis of the Tiberian Hebrew System of Accents. *Language* 70(1). 1-52.

Duanmu, San. 1992. A Featural Analysis of Some Onset-vowel Interactions. *Papers from the First Annual Meeting of the Southeast Asian Linguistics Society*, Arizona State University. 141-158.

Duanmu, San. 1995. Metrical and Tonal Phonology of Compounds in Two Chinese Dialects. *Language* 71(2). 225-259.

Duanmu, San. 1998. Wordhood in Chinese. *New Approaches to Chinese Word Formation: Morphology, Phonoloy and the Lexicon in Modern and Ancient Chinese*, ed. by Jerome Packard, 135-196. Berlin/New York: Mouton de Gruyter.

Duanmu, San. 2005. The tone-syntax interface in Chinese: some recent controversies. *Proceedings of the Symposium "Cross-Linguistic Studies of Tonal Phenomena, Historical Development, Tone-Syntax Interface, and Descriptive Studies"*, Tokyo. 221-254.

Duanmu, San. 2007. *The phonology of Standard Chinese. Second edition.* Oxford: Oxford University Press.

Duanmu, San. 2012. Word-length preferences in Chinese: a corpus study. *Journal of East Asian Linguistics* 21(1). 89-114.

É. Kiss, Katalin. 1998. Identificational focus versus information focus. *Language* 74(2). 245-273.

É. Kiss, Katalin, Ed. 2009. *Adverbs and Adverbial Adjuncts at the Interfaces*. Interface Explorations. Berlin/New York: Mouton de Gruyter.

Eady, Stephen J., William E. Cooper, Gayle V. Klouda, Pamela R. Mueller, and Dan W. Lotts. 1986. Acoustical Characteristics of Sentential Focus: Narrow vs. Broad and Single vs. Dual Focus Environments. *Language and Speech* 29(3). 233-251.

Féry, Caroline. 2010. Syntax, information structure, embedded prosodic phrasing, and the relational scaling of pitch accents. *The Sound Patterns of Syntax*, ed. by Nomi Erteschik-Shir and Lisa Rochman, 271-290. Oxford: Oxford University Press.

Féry, Caroline, and Shinichiro Ishihara. 2009. The Phonology of Second Occurrence Focus. *Journal of Linguistics* 45(2). 285-313.

Féry, Caroline, and Shinichiro Ishihara. 2010. How focus and givenness shapes prosody. *Information structure from different perspectives*, ed. by Malte Zimmermann and Caroline Féry, 36-63. Oxford: Oxford University Press.

Féry, Caroline, and Frank Kügler. 2008. Pitch accent scaling on given, new and focused constituents in German. *Journal of Phonetics* 36(4). 680-703.

Fougeron, Cécile. 1999. Prosodically conditioned articulatory variations: A review. *UCLA Working Papers in Phonetics* 97(1). 1-73.

Fougeron, Cécile, and Patricia A. Keating. 1997. Articulatory strengthening at edges of prosodic domains. *Journal of the Acoustic Society of America* 101(6). 3728-3740.

Frascarelli, Mara. 1999. The prosody of Focus in Italian (and the syntax-phonology interface). *Probus* 11(2). 209-238.

Frascarelli, Mara. 2000. *The Syntax-Phonology Interface in Focus and Topic Constructions in Italian*. Studies in Natural Language and Linguistic Theory 50. Dordrecht: Kluwer.

Frota, Sonia. 2000. *Prosody and Focus in European Portugese*. New York: Garland.

Fu, Guotong, and Songxi Fang. 1985. *浙江吴语分区 [Sub-groups of the Wu dialects of Zhejiang]*. Zhoushan: 浙江省方言研究会 [Zhejiang dialect study research association].

Gandour, Jackson, Siripong Potisuk, and Sumalee Dechongkit. 1994. Tonal coarticulation in Thai. *Journal of Phonetics* 22(4). 477-492.

Gårding, Eva. 1979. Sentence Intonation in Swedish. *Phonetica* 36(3). 207-215.

Gu, Wentao, and Tan Lee. 2007a. Effects of Focus on Prosody of Cantonese Speech – A Comparison of Surface Feature Analysis and Model-Based Analysis. *ParaLing 07: Proceedings of the International Workshop on Paralinguistic Speech - between Models and Data*, Saarbrücken. 59-64.

Gu, Wentao, and Tan Lee. 2007b. Effects of tonal context and focus on Cantonese F0. *Proceedings of the 16th International Congress of Phonetic Sciences*, Saarbrücken. 1033-1036.

Gundel, Jeanette K., and Thorsten Fretheim. 2006. Topic and Focus. *The Handbook of Pragmatics*, ed. by Laurence R. Horn and Gregory Ward, 175-196. Oxford: Blackwell.

Gundel, Jeanette K., Nancy Hedberg, and Ron Zacharski. 1993. Cognitive Status and the Form of Referring Expressions in Discourse. *Language* 69(2). 274-307.

Gussenhoven, Carlos. 1992. Sentence accents and argument structure. *Thematic structure: its role in grammar*, ed. by Iggy M. Roca, 79-106. Berlin: Foris.

Gussenhoven, Carlos. 2007. Types of Focus in English. *Topic and Focus: Cross-Linguistic Perspectives on Meaning and Intonation*, ed. by Chungmin Lee, Matthew Gordon and Daniel Büring, 83-100. Dordrecht: Springer.

Halliday, Michael A. K. 1967. Notes in transitivity and theme in English - Part 2. *Journal of Linguistics* 3(2). 199-244.

Han, Mieko S., and Kong-On Kim. 1974. Phonetic variation of Vietnamese tones in disyllabic utterances. *Journal of Phonetics* 2(2). 223-232.

Hanssen, Judith, Jörg Peters, and Carlos Gussenhoven. 2008. Prosodic Effects of Focus in Dutch Declaratives. *Proceedings of the 4th Conference on Speech Prosody*, Campinas. 609-612.

Hayashi, Midori. 2004. Downstep and prosodic structure in Japanese. *Toronto Working Papers in Linguistics* 23(2). 47-70.

Hayashi, Wendy, Chai-Shune Hsu, and Patricia Keating. 1999. Domain-initial strengthening in Taiwanese: a follow-up study. *UCLA Working Papers in Phonetics* 97(1). 152-156.

Hayes, Bruce. 1995. *Metrical Stress Theory*. Chicago: University of Chicago Press.

Hayes, Bruce, and Aditi Lahiri. 1991. Bengali Intonational Phonology. *Natural Language and Linguistic Theory* 9(1). 47-96.

Hayward, Katrina, Justin Watkins, and Akin Oyétádé. 2003. The phonetic interpretation of register: evidence from Yorúbá. *Papers in Laboratory Phonology VI: Phonetic Interpretation*, ed. by John Local, Richard Ogden and Rosalind Temple, 305-321. Cambridge: Cambridge University Press.

Hellmuth, Sam. 2007. The relationship between prosodic structure and pitch accent distribution: evidence from Egyptian Arabic. *The Linguistic Review* 24(2-3). 291-316.

Hou, Jingyi. 1998. *温州话音档 [Wenzhou speech sounds]*. 现代汉语方言音库 [Modern Chinese Dialect Sound Database]. Shanghai: 上海教育出版社 [Shanghai educational press].

Hsu, Chai-Shune K., and Sun-Ah Jun. 1998. Prosodic Strengthening in Taiwanese: Syntagmatic or Paradigmatic? *UCLA Working Papers in Phonetics* 96(1). 69-89.

Huang, C.-T. James. 1984. Phrase Structure, Lexical Integrity and Chinese Compounds. *Journal of the Chinese Language Teachers Association* 19(1). 53-78.

Huang, C.-T. James. 1987. Remarks on Empty Categories in Chinese. *Linguistic Inquiry* 18(2). 321-337.

Ishihara, Shinichiro. 2007. Intonation of Sentences with an NPI. *Proceedings of the 2nd Workshop on Prosody, Syntax and Information Structure (WPSI2)*, Potsdam. 65-96.

Ishihara, Shinichiro. 2011. Japanese focus prosody revisited: Freeing focus from prosodic phrasing. *Lingua* 121(13). 1870-1889.

Jackendoff, Ray. 1972. *Semantic Interpretation in Generative Grammar*. Cambridge, MA: MIT Press.

Jiang, Ping, and Aishu Chen. 2011. Representation of Mandarin intonations: boundary tone revisited. *Proceedings of the 23rd North American Conference on Chinese Linguistics*, Eugene. 97-109.

Jin, Shunde. 1996. An Acoustic Study of Sentence Stress in Mandarin Chinese. Ohio: Ohio State University PhD dissertation.

Jun, Sun-Ah. 1998. The Accentual Phrase in the Korean prosodic hierarchy. *Phonology* 15(2). 189-226.

Jun, Sun-Ah, and Hee-Sun Kim. 2007. VP Focus and Narrow Focus in Korean. *Proceedings of the 16th International Congress of Phonetic Sciences*, Saarbrücken. 1277-1280.

Kabagema-Bilan, Elena, Beatriz López-Jiménez, and Hubert Truckenbrodt. 2011. Multiple focus in Mandarin Chinese. *Lingua* 121(13). 1890-1905.

Kahnemuyipour, Arsalan. 2003. Syntactic Categories and Persian Stress. *Natural Language and Linguistic Theory* 21(2). 333-379.

Katz, Jonah, and Elisabeth Selkirk. 2011. Contrastive Focus vs. Discourse-New: Evidence from Phonetic Prominence in English. *Language* 87(4). 771-816.

Kawahara, Shigeto, and Takahito Shinya. 2008. The Intonation of Gapping and Coordination in Japanese: Evidence for Intonational Phrase and Utterance. *Phonetica* 65(1-2). 62-105.

Keating, Patricia, Taehong Cho, Cécile Fougeron, and Chai-Shune Hsu. 2003. Domain-initial articulatory strengthening in four languages. *Papers in Laboratory Phonology VI: Phonetic Interpretation*, ed. by Richard Ogden & Rosalind Temple John Local, 145-163. Cambridge: Cambridge University Press.

Krifka, Manfred. 2007. Basic Notions of Information Structure. *The Notions of Information Structure. Interdisciplinary Studies on Information Structure (ISIS) 6*, ed. by Caroline Féry, Gisbert Fanselow and Manfred Krifka, 13-55. Potsdam: Potsdam University Press.

Kubozono, Haruo. 1993. *The organization of Japanese prosody*. Tokyo: Kurosio Publishers.

Kügler, Frank. 2008. The role of duration as phonetic correlate of focus. *Proceedings of the 4th Conference on Speech Prosody*, Campinas. 591-594.

Ladd, D. Robert. 1996. *Intonational Phonology*. Cambridge: Cambridge University Press.

Ladd, D. Robert. 2008. *Intonational Phonology. Second Edition*. Cambridge: Cambridge University Press.

Ladd, D. Robert, and Catherine Johnson. 1987. "Metrical" factors in the scaling of sentence-initial accent peaks. *Phonetica* 44(4). 238-245.

Ladefoged, Peter, and Ian Maddieson. 1996. *The Sounds of the World's Languages*. Oxford: Blackwell.

Lambrecht, Knud. 1994. *Information structure and sentence form: Topic, focus, and the mental representation of discourse referents*. Cambridge Studies in Linguistics 71. Cambridge: Cambridge University Press.

Laniran, Yetunde O., and G.N. Clements. 2003. Downstep and high raising: interacting factors in Yoruba tone production. *Journal of Phonetics* 31(2). 203-250.

Lee, Kent A. 2002. Chinese Tone Sandhi and Prosody. Urbana: University of Illinois at Urbana-Campaign PhD dissertation.

Lee, Wai-Sum, and Eric Zee. 2003. Standard Chinese (Beijing). *Journal of the International Phonetic Association* 33(1). 109-112.

Li, Charles N., and Sandra A. Thompson. 1981. *Mandarin Chinese. A Functional Reference Grammar*. Berkeley: University of California Press.

Li, Xingjian, and Sixun Liu. 1985. 天津方言的连读变调 [Tone sandhi in connected speech in the Tianjin dialect]. *中国语文 [Chinese language]* 1(1). 76-80.

Li, Xuping, and Walter Bisang. 2012. Classifiers in Sinitic languages: From individuation to definiteness-marking. *Lingua* 122(4). 335-355.

Liberman, Mark, and Janet Pierrehumbert. 1984. Intonational invariance under changes in pitch range and length. *Language sound structure: Studies in phonology presented to Morris Halle by his teacher and students*, ed. by Mark Aronoff and Richard T Oehrle, 157-233. Cambridge, MA: MIT Press.

Lin, Hui-shan. 2005. Prosodic Correspondence in Tone Sandhi. *UST Working Papers in Linguistics* 1(1). 229-265.

Lin, Hwei-Bing. 1988. Contextual stability of Taiwanese tones. University of Connecticut PhD dissertation.

Lin, Jowang. 1994. Lexical government and tone group formation in Xiamen Chinese. *Phonology* 11(2). 237-276.

Liu, Fang, and Yi Xu. 2005. Parallel Encoding of Focus and Interrogative Meaning in Mandarin Intonation. *Phonetica* 62(1). 70-87.

Nagahara, Hiroyuki. 1994. Phonological phrasing in Japanese. Los Angeles: University of California Los Angeles PhD dissertation.

Nakajima, Motoki. 1983. *浙南 吴语 基础 语汇集. [The Southern Chekiang dialect: a linguistic study of its phonology and lexicon]*. Tokyo: 东京外国语大学アジア・アフリカ语言文化研究所 [Institute for the Study of Languages and Cultures of Asia and Africa].

Nespor, Marina, and Irene Vogel. 1986. *Prosodic Phonology*. Dordrecht: Kluwer.

Nespor, Marina, and Irene Vogel. 2007. *Prosodic Phonology: with a new foreword*. Studies in Generative Grammar 28. Berlin/New York: Mouton de Gruyter.

Norman, Jerry. 1988. *Chinese*. Cambridge: Cambridge University Press.

Pan, Ho-Hsien. 2007a. The effect of prosodic boundaries on nasality in Taiwan Min. *Journal of the Acoustic Society of America* 121(6). 3755-3769.

Pan, Ho-Hsien. 2007b. Focus and Taiwanese Unchecked Tones. *Topic and Focus. Cross-Linguistic Perspectives on Meaning and Intonation*, ed. by Chungmin Lee, Matthew Gordon and Daniel Büring, 195-213. Dordrecht: Springer.

Pan, Ho-Hsien, Wan-Ting Huang, and Ying-Hui Huang. 2005. Production and Perception of New, Given, and Contrastive Information in Taiwan Mandarin. *Proceedings of the ISCA Workshop on Plasticity in Speech Perception (PSP 2005)*, London. 87-90.

Pan, Ho-Hsien, and Yi-hsin Tai. 2006. Boundaries and Tonal Articulation in Taiwanese Min. *Proceedings of the 3rd International Conference on Speech Prosody*, Dresden. 51-54.

Pan, Wuyun. 1991. An Introduction to the Wu dialects. *Languages and Dialects of China*, ed. by William S-Y. Wang, 237-293. Berkeley: Project on Linguistic Analysis.

Peng, Shu-Hui. 1997. Production and perception of Taiwanese tones in different tonal and prosodic contexts. *Journal of Phonetics* 25(3). 371-400.

Peng, Shu-Hui. 2000. Lexical versus 'phonological' representations of Mandarin sandhi tones. *Papers in Laboratory Phonology V: Acquisition and the Lexicon*, ed. by Michael B. Broe and Janet B. Pierrehumbert, 152-167. Cambridge: Cambridge University Press.

Pierrehumbert, Janet. 1979. The perception of fundamental frequency declination. *Journal of the Acoustic Society of America* 66(2). 363-369.

Pierrehumbert, Janet. 1980. The Phonology and Phonetics of English Intonation. Cambridge, MA: MIT PhD dissertation.

Pierrehumbert, Janet, and Mary Beckman. 1988. *Japanese Tone Structure*. Cambridge, MA: MIT Press.

Pierrehumbert, Janet, and Julia Hirschberg. 1990. The Meaning of Intonational Contours in the Interpretation of Discourse. *Intentions in Communication*, ed. by Philip R. Cohen, Jerry Morgan and Martha E. Pollack, 271-311. Cambridge, MA: MIT Press.

Potisuk, Siripong, Jack Gandour, and Mary P. Harper. 1997. Contextual Variations in Trisyllabic Sequences of Thai Tones. *Phonetica* 54(1). 22-42.

Prieto, Pilar, Mariapaola D'Imperio, Gorka Elordieta, Sónia Frota, and Marina Vigário. 2006. Evidence for soft preplanning in tonal production: Initial scaling in Romance. *Proceedings of the 3rd International Conference on Speech Prosody*, Dresden. 803-806.

Prieto, Pilar, Chilin Shih, and Holly Nibert. 1996. Pitch downtrend in Spanish. *Journal of Phonetics* 24(4). 445-473.

Qian, Nairong. 1992. *当代吴语研究 [Studies in the Contemporary Wu Dialects]*. Shanghai: 上海教育出版社 [Shanghai educational press].

Rialland, Annie. 2001. Anticipatory raising in downstep realization:evidence for preplanning in tone production. *Proceedings of the Symposium on Cross-linguistics Studies of Tonal Phenomenon: Tonogenesis, Typology, and Related Topics*, Tokyo. 301-322.

Rose, Phil. 2000. Wenzhou dialect disyllabic lexical tone sandhi with first syllable entering tones. *Proceedings of the 8th Australian International Conference on Speech Science and Technology*, Canberra. 230-235.

Rose, Phil. 2001. Independent Depressor and Register Effects in Wu Dialect Tonology: Evidence from Wenzhou Tone Sandhi. *Journal of Chinese Linguistics* 30(1). 39-81.

Rose, Phil. 2002. Tonal Complexity as Conditioning Factor: More Depressing Wenzhou Dialect Disyllabic Lexical Tone Sandhi. *Proceedings of the 9th Australian International Conference on Speech Science and Technology*, Canberra. 64-69.

Rose, Phil. 2004. „Defying Explanation?"- Accounting for Tones in Wenzhou Dialect Disyllabic Lexical Tone Sandhi. *Proceedings of the 10th Australian International Conference on Speech Science and Technology*, Canberra. 237-242.

Rose, Phil. 2008. Oujiang Wu tones and acoustic reconstruction. *Morphology and language history: in honour of Harold Koch*, ed. by Claire Bowern, Bethwyn Evans and Luisa Miceli, 235-250. Amsterdam: John Benjamins.

Samek-Lodovici, Vieri. 2005. Prosody-Syntax Interaction in the Expression of Focus. *Natural Language and Linguistic Theory* 23(3). 687-755.

Schwarzschild, Roger. 1999. GIVENness, AvoidF and other constraints on the placement of accent. *Natural Language Semantics* 7(2). 141-177.

Selkirk, Elisabeth. 1984. *Phonology and syntax: The relation between sound and structure*. Cambridge, MA: MIT Press.

Selkirk, Elisabeth. 1986. On derived domains in sentence phonology. *Phonology Yearbook* 3. 371-405.

Selkirk, Elisabeth. 2007a. Bengali Intonation Revisited: An Optimality Theoretic Analysis in which FOCUS Stress Prominence Drives FOCUS Phrasing. *Topic and Focus. Cross-Linguistic Perspectives on Meaning and Intonation*, ed. by Chungmin Lee, Matthew Gordon and Daniel Büring, 215-244. Dordrecht: Springer.

Selkirk, Elisabeth. 2007b. Contrastive Focus, Givenness and the Unmarked Status of "Discourse New". *The Notions of Information Structure. Interdisciplinary Studies on Information Structure (ISIS) 6*, ed. by Caroline Féry, Gisbert Fanselow and Manfred Krifka, 125-145. Potsdam: Potsdam University Press.

Selkirk, Elisabeth, and Tong Shen. 1990. Prosodic Domains in Shanghai Chinese. *The Phonology-Syntax Connection*, ed. by Sharon Inkelas and Draga Zec, 313-337. Chicago/London: The University of Chicago Press.

Selkirk, Elisabeth, and Koichi Tateishi. 1991. Syntax and downstep in Japanese. *Interdisciplinary approaches to language: Essays in honor of S.-Y. Kuroda*, ed. by Carol Georgopoulos and Roberta Ishihara, 519-543. Dordrecht: Kluwer Academic Publishers.

Shen, Xiaonan Susan. 1990a. *The Prosody of Mandarin Chinese*. University of California Publications in Linguistics 118. Berkeley: University of California Press.

Shen, Xiaonan Susan. 1990b. Tonal coarticulation in Mandarin. *Journal of Phonetics* 18(2). 281-295.

Shen, Xiaonan Susan. 1992a. On Tone Sandhi and Tonal Coarticulation. *Acta Linguistica Hafniensia* 25(1). 83-94.

Shen, Xiaonan Susan. 1992b. A Pilot Study on the Relation between the Temporal and Syntactic Structures in Mandarin. *Journal of the International Phonetic Association* 22(1-2). 35-43.

Shen, Xiaonan Susan, and Maocan Lin. 1991. Concept of Tone in Mandarin Revisited: A Perceptual Study on Tonal Coarticulation. *Language Sciences* 13(3-4). 421-432.

Shi, Feng. 1983. 苏州话浊塞音的声学特征 [The acoustic characteristics of the voiced stops in Suzhou dialect]. *语言研究 [Studies in Language and Linguistics]* 3(1). 49-83.

Shih, Chilin. 1986. The Prosodic Domain of Tone Sandhi in Chinese. La Jolla: University of California San Diego PhD dissertation.

Shih, Chilin. 1988. Tone and Intonation in Mandarin. *Working Papers of the Cornell Phonetics Laboratory* 3(1). 83-109.

Shih, Chilin. 1997. Mandarin third tone sandhi and prosodic structure. *Studies in Chinese Phonology*, ed. by Wang Jialing and Norval Smith, 81-123. Berlin/New York: Mouton de Gruyter.

Shih, Chilin. 2000. A Declination Model of Mandarin Chinese. *Intonation. Analysis, Modelling and Technology*, ed. by Antonis Botinis, 243-268. Dordrecht: Kluwer Academic Publishers.

Shih, Chilin. 2004. Tonal Effects on Intonation. *Proceedings of the International Symposium on Tonal Aspects of Languages: With Emphasis on Tone Languages*, Beijing. 163-168.

Skopeteas, Stavros, Ines Fiedler, Sam Hellmuth, Anne Schwarz, Ruben Stoel, Gisbert Fanselow, Caroline Féry, and Manfred Krifka. 2006. *Questionnaire on Information Structure (QUIS): Reference Manual*. ISIS Working Papers of the SFB 632 ; 4. Potsdam: Potsdam University Press.

Sluijter, Agaath. 1992. Lexical stress and focus distribution as determinants of temporal structure. *Linguistics in the Netherlands*, ed. by Reineke Bok-Bennema and Roeland van Hout, 247-259. Amsterdam: John Benjamins.

Sluijter, Agaath, and Vincent J. van Heuven. 1995. Effects of Focus Distribution, Pitch Accent and Lexical Stress on the Temporal Organization of Syllables in Dutch. *Phonetica* 52(2). 71-89.

Soh, Hooi Ling. 2001. The syntax and semantics of phonological phrasing in Shanghai and Hokkien. *Journal of East Asian Linguistics* 10(1). 37-80.

Sorensen, John M., and William E. Cooper. 1980. Syntactic coding of fundamental frequency in speech production. *Perception and Production of Fluent Speech*, ed. by Ronald A.Cole, 399-440. Hillsdale: Lawrence Erlbaum Associates.

Sugahara, Mariko. 2003. Downtrends and post-FOCUS intonation in Japanese. Amherst, MA: University of Massachussetts PhD dissertation.

Tang, Chaoju, and Vincent J. van Heuven. 2007. Mutual intelligibility and similarity of Chinese dialects. Predicting judgements from objective measures. *Linguistics in the Netherlands 2007*, ed. by Bettelou Los and Marjo van Koppen, 223-234. Amsterdam: Benjamins.

Tang, Chaoju, and Vincent J. van Heuven. 2009. Mutual intelligibility of Chinese dialects experimentally tested. *Lingua* 119(5). 709-732.

Thorsen, Nina. 1980. Intonation contours and stress group patterns in declarative sentences of varying length in ASC Danish. *Annual report of the Institute of Phonetics* 15(1). 13-47.

Truckenbrodt, Hubert. 1995. Phonological Phrases: Their Relation to Syntax, Focus and Prominence. Boston: MIT PhD dissertation.

Truckenbrodt, Hubert. 1999. On the Relation between Syntactic Phrases and Phonological Phrases. *Linguistic Inquiry* 30(2). 219-255.

Truckenbrodt, Hubert. 2002. Upstep and embedded register levels. *Phonology* 17(1). 77-120.

Truckenbrodt, Hubert. 2007. Upstep of edge tones and of nuclear accents. *Tones and Tunes. Volume 2: Experimental Studies in Word and Sentence Prosody*, ed. by Carlos Gussenhoven and Thomas Riad, 349-386. Berlin: Mouton.

Truckenbrodt, Hubert, and Caroline Féry. 2003. More on hierarchical organization and tonal scaling. Tübingen/Potsdam, ms.

Tsay, Jane, and James Myers. 1996. Taiwanese tone sandhi as allomorph selection. *Proceedings of the 22nd Annual Meeting of the Berkeley Linguistic Society*, Berkeley. 394-405.

Tsay, Jane, James Myers, and Xiao-Jun Chen. 1999. Tone Sandhi as Evidence for Segmentation in Taiwanese. *Proceedings of the 30th Annual Child Language Research Forum*, Stanford. 211-218.

van den Berg, Rob, Carlos Gussenhoven, and Toni Rietveld. 1992. Downstep in Dutch: Implications for a model. *Papers in Laboratory Phonology II: Gesture, Segment, Prosody*, ed. by Gerard J. Docherty and D. Robert Ladd, 335-367. Cambridge: Cambridge University Press.

van Heuven, Vincent J. 1994. What is the smallest prosodic domain? *Papers in Laboratory Phonology III: Phonological structure and phonetic form*, ed. by Patricia A. Keating, 76-98. Cambridge: Cambridge University Press.

Wang, Bei, and Yi Xu. 2006. Prosodic Encoding of Topic and Focus in Mandarin. *Proceedings of the 3rd International Conference on Speech Prosody*, Dresden. PS3-12.

Wang, Bei, and Yi Xu. 2011. Differential prosodic encoding of topic and focus in sentence-initial position in Mandarin Chinese. *Journal of Phonetics* 39(4). 595-611.

Wang, H. Samuel. 2002. The Prosodic Effects on Taiwan Min Tones. *Language and Linguistics* 3(4). 839-852.

Watson, Duane, and Edward Gibson. 2004. The relationship between intonational phrasing and syntactic structure in language production. *Language and Cognitive Processes* 19(6). 713-755.

Wei, Yang. 1997. On the Control Structure of the Chinese Verb Daying. *Working Papers of the Linguistics Circle of the University of Victoria* 14(1). 89-105.

Whalen, D.H., and Andrea G. Levitt. 1995. The universality of intrinsic F0 of vowels. *Journal of Phonetics* 23(3). 349-366.

Wu, Wing Li, and Yi Xu. 2010. Prosodic Focus in Hong Kong Cantonese without Post-focus Compression. *Proceedings of the 5th International Conference on Speech Prosody*, Chicago.

Xu, Liejiong. 2004. Manifestation of informational focus. *Lingua* 114(3). 277-299.

Xu, Yi. 1994. Production and perception of coarticulated tones. *Journal of the Acoustic Society of America* 95(4). 2240-2253.

Xu, Yi. 1997. Contextual tonal variations in Mandarin. *Journal of Phonetics* 25(1). 61-83.

Xu, Yi. 1999. Effects of tone and focus on the formation and alignment of f0 contours. *Journal of Phonetics* 27(1). 55-105.

Xu, Yi. 2004. Separation of functional components of tone and intonation from observed F0 patterns. *From Traditional Phonology to Modern Speech Processing: Festschrift for Professor Wu Zongji's 95th Birthday*, ed. by Gunnar Fant, Hiroya Fujisaki, Jianfen Cao and Yi Xu, 483-505. Beijing: Foreign Language Teaching and Research Press.

Xu, Yi. 2005. Speech melody as articulatorily implemented communicative functions. *Speech Communication* 46(3-4). 220-251.

Xu, Yi, Szu-wei Chen, and Bei Wang. 2012. Prosodic focus with and without post-focus compression: A typological divide within the same language family? *The Linguistic Review* 29(1). 131-147.

Xu, Yi, and Xuejing Sun. 2002. Maximum speed of pitch change and how it may relate to speech. *Journal of the Acoustic Society of America* 111(3). 1399-1413.

Xu, Yi, and Maolin Wang. 2009. Organizing syllables into groups - Evidence from F0 and duration patterns in Mandarin. *Journal of Phonetics* 37(4). 502-520.

Xu, Yi, and Q. Emily Wang. 2001. Pitch targets and their realization: Evidence from Mandarin Chinese. *Speech Communication* 33(4). 319-337.

Xu, Yi, Ching X. Xu, and Xuejing Sun. 2004. On the Temporal Domain of Focus. *Proceedings of the 2nd International Conference on Speech Prosody*, Nara. 81-84.

Yang, Yufang, and Bei Wang. 2002. Acoustic Correlates of Hierarchical Prosodic Boundary in Mandarin. *Proceedings of the 1st International Conference on Speech Prosody*, Aix-en-Provence. 707-710.

Yip, Moira. 1993. Tonal Register in East Asian Languages. *The Phonology of Tone. The Representation of Tonal Register*, ed. by Harry van der Hulst and Keith Snider, 245-268. Berlin/New York: Mouton de Gruyter.

Yip, Moira. 1999. Feet, tonal reduction and speech rate at the word and phrase level in Chinese. *Phrasal Phonology*, ed. by René Kager and Wim Zonneveld, 171-194. Nijmegen: Nijmegen University Press.

You, Rujie, and Qianming Yang. 1998. *温州方言词典 [Wenzhou dialect dictionary]*. 方言词典 [Dialect dictionaries]. Nanjing: 江苏教育出版社 [Jiangsu educational press].

You, Rujue. 2002. 温州方言语法纲要 [A grammar of the Wenzhou dialect]. *著名中年语言学家自选集. 游汝杰卷 [Famous Chinese middle-age linguists series: You Rujie volume]*, ed. by Zhichun Wan, 164-235. Hefei: 安徽教育出版社 [Anhui educational press].

Yuan, Jiahong. 2004. Intonation in Mandarin Chinese: Acoustics, Perception, and Computational Modeling. Ithaca: Cornell University PhD dissertation.

Yue-Hashimoto, Anne. 1987. Tone sandhi across Chinese dialects. *王力先生纪念论文集 [Wang Li Memorial Volumes: English Volume]*, ed. by, 445-474. Hong Kong: Chinese Language Society of Hong Kong.

Zec, Draga, and Sharon Inkelas. 1990. Prosodically Constrained Syntax. *The Phonology-Syntax Connection*, ed. by Sharon Inkelas and Draga Zec, 365-378. Chicago: University of Chicago Press.

Zhang, Jie. 2007a. Contour tone distribution is not an artifact of tonal melody mapping. *Studies in the Linguistic Sciences* 33(1). 1-61.

Zhang, Jie. 2007b. A directional asymmetry in Chinese tone sandhi systems. *Journal of East Asian Linguistics* 16(4). 259-302.

Zhang, Jie, and Yuwen Lai. 2006. Testing the Role of Phonetic Naturalness in Mandarin Tone Sandhi. *Kansas Working Papers in Linguistics* 28(1). 65-126.

Zhang, Jie, and Yuwen Lai. 2010. Testing the role of phonetic knowledge in Mandarin tone sandhi. *Phonology* 27(1). 153-201.

Zhang, Jie, and Jiang Liu. 2011. Tone Sandhi and Tonal Coarticulation in Tianjin Chinese. *Phonetica* 68(4). 161-191.

Zhang, Jiesheng. 2006. The Phonology of Shaoxing Chinese. Leiden: PhD dissertation.

Zhang, Ning. 1997. The Avoidance of the Third Tone Sandhi in Mandarin Chinese. *Journal of East Asian Linguistics* 6(4). 293-338.

Zheng, Yuling, Jianfen Cao, and Huaiqiao Bao. 2006. Coarticulation and prosodic hierarchy. Beijing: Chinese Academy of Social Sciences, ms.

Zhengzhang, Shangfang. 1964a. 温州方言的连读变调 [Tone Sandhi in the Wenzhou Dialect]. *中国语文 [Chinese language]* 129(1). 106-152.

Zhengzhang, Shangfang. 1964b. 温州音系 [Wenzhou sound system]. *中国语文 [Chinese language]* 129(1). 28-60.

Zhengzhang, Shangfang. 2008. *温洲方言志 [Record of the Wenzhou dialect]*. Beijing: 中华书局 [China publishing house].

Samenvatting in het Nederlands

Een van de spannendste onderdelen van taalwetenschap is het onderzoek naar de spreker-luisteraar-verhouding en hoe sprekers informatie voor luisteraars coderen. Het is al lang bekend dat sprekers in veel talen bijzondere middelen gebruiken om een of meerdere woorden in een zin te benadrukken en er de aandacht van de luisteraar op te richten. Zo passen sprekers bijvoorbeeld hun stem aan om een belangrijk woord in een zin luider, hoger en/of langer te laten klinken dan gewoonlijk. Luisteraars op hun beurt pikken die informatie op en beseffen dat het benadrukte woord een belangrijke rol speelt in de conversatie.

In sommige talen, zoals de meeste dialecten van het Chinees, zijn dergelijke middelen beperkt: In deze talen wordt "toon", d.w.z. de hoogte (hoog-laag) en verandering (stijging-daling) van de stemtoon, gebruikt om woorden in betekenis te laten verschillen. Als bijvoorbeeld in het Standaard Chinees de lettergreep *ma* met een hoge toon wordt uitgesproken, betekent hij "moeder", maar dezelfde lettergreep met een lage toon betekent "paard".

Dit proefschrift gaat over een zuidelijk dialect van het Chinees, met name het dialect dat in de stad Wenzhou wordt gesproken. De vraag is: Hoe benadrukken sprekers van dit dialect een woord zonder de betekenisverschillen uit te wissen? Om die vraag te beantwoorden werden er een aantal geluidsopnamen gemaakt van sprekers van dit dialect en werd onderzocht hoe zij nadruk leggen onder verschillende condities.

Na een inleiding, waarin eerder onderzoek naar de verschillende onderdelen van dit proefschrift wordt samengevat, geeft het tweede hoofdstuk een overzicht van de klinkers en woordtonen van het Wenzhou dialect. Vervolgens geeft dit hoofdstuk een inleiding op een verschijnsel dat vaak onder de naam "toonsandhi" onderzocht wordt: Als bijvoorbeeld twee lettergrepen onmiddellijk na elkaar worden uitgesproken, verandert vaak de toon op beide lettergrepen. Dit kan gebeuren omdat de spreker wil aangeven dat de twee lettergrepen bij elkaar horen, bijvoorbeeld in een samenstelling zoals *hoofdpijn* in de zin *Ik voel hoofdpijn*. Dit woord in het Wenzhou klinkt anders dan wanneer de spreker de woorden voor *hoofd* en *pijn* als aparte woorden uitspreekt in een zin als *Ik voel dat mijn hoofd pijn doet*. Door de toonverandering laat de spreker dus de luisteraar merken dat hij het samengestelde woord bedoelt, en niet de twee woorden los van elkaar.

Een belangrijke vraag, die in hoofdstuk drie onderzocht wordt, is wat er met samenstellingen gebeurt die qua grammatica niet echt tot de samenstellingen worden gerekend, maar waarvan de delen ook niet echt onafhankelijk van elkaar bekeken kunnen worden. Het gaat bijvoorbeeld om samenstellingen zoals *auto rijden*, waarin de betekenis van *rijden* anders is dan bijvoorbeeld in de samenstelling *paard rijden*. Je kunt dus zeggen dat de context van *auto* of *paard* bepaalt wat precies de betekenis van *rijden* is, maar tegelijkertijd bevatten beide samenstellingen een werkwoord (verbum = V) en een voorwerp (object = O), en zijn dus duidelijk uit twee woorden samengesteld. Welke criteria bepalen of sprekers voor dergelijke VO-constructies een samenstellingstooncontour gebruiken, of dat ze de twee delen als aparte woorden uitspreken?

De experimenten die in hoofdstuk drie beschreven worden, laten zien dat de mate waarin de twee woorden samenhoren een belangrijke rol speelt in de keuze van de sprekers voor of tegen een samenstellingscontour. Sprekers van het Wenzhou gebruiken liever de vorm met samenstellingscontour als de twee onderdelen van de constructie meer samen horen, bijvoorbeeld in een samenstelling waarvan de betekenis niet gemakkelijk afleidbaar is uit de betekenissen van de delen.

In tegenstelling tot eerdere onderzoeken naar dit soort constructies laat het hoofdstuk ook zien dat focus op de constructie nauwelijks invloed heeft op keuze van tooncontour. De belangrijkste factor voor verandering van tooncontour in VO-constructies is de context: Als sprekers een lijst met alleen de variabele VO-constructies moeten lezen, realiseren ze de twee woorden van elke constructie meestal apart. Als de variable VO-constructies daarentegen in een lijst samen met echte samenstellingen (zoals *hoofdpijn*) voorkomen, gebruiken de sprekers meestal de samenstellingscontour, ook voor VO-constructies waarin de delen aparte woorden zijn.

Ook in het vierde hoofdstuk gaat het om die toonverandering, maar deze keer alleen maar in echte samenstellingen. De vraag is: Wat gebeurt er met de toonverandering als de spreker alleen maar een van de twee lettergrepen in de samenstelling wil benadrukken, zoals in de zin *Ik zei niet spierpijn maar hoofdpijn* of *Ik zei niet hoofdhaar maar hoofdpijn*? Hij geeft dus met de toonverandering aan dat de twee lettergrepen qua betekenis bij elkaar horen, maar tegelijkertijd moet hij ook duidelijk maken dat er een contrast bestaat, en op welke van de twee lettergrepen dit contrast te vinden is. Contrast op de eerste of tweede lettergreep wordt vergeleken met contrast op het hele woord, en ook met contrast onmiddellijk voor of na het doelwoord.

De resultaten in hoofdstuk vier laten zien dat er in het Wenzhou dialect inderdaad nauwelijks een verschil is in de manier waarop de toonverandering uitgevoerd wordt als er een contrast speelt op de eerste lettergroep, de tweede lettergroep, of het hele woord. Dat laat zien dat het voor de spreker belangrijker is om de samenstelling als één geheel te realiseren, ook al kan hij dan niet precies aangeven waar het contrast zit. Wel is er een verschil tussen de condities waarbij het contrast binnen het doelwoord ligt, en de condities waarbij het contrast voor of na het doelwoord ligt (dus op een eerder of later woord). Contrast ergens binnen het doelwoord veroorzaakt dat het hele doelwoord met een grotere toonbeweging en een langere duur wordt uitgesproken. Dit laat zien dat de sprekers wel in staat zijn om nadruk ook in de toonveranderde woorden uit te drukken, alleen dan binnen de beperkingen van de toonverandering.

Ook in hoofdstuk vijf gaat het om toonverandering, maar deze keer om een andere soort toonverandering, namelijk tooncoarticulatie. Het gaat om woorden die maar uit één lettergreep bestaan en een stijgende of dalende tooncontour hebben. Als dit soort woorden in de context van andere woorden wordt uitgesproken, verandert de grootte en/of de steilheid van de stijging of daling. Als bijvoorbeeld op een stijgende toon weer een stijgende toon volgt, wordt de stijging op de eerste toon niet compleet uitgevoerd, omdat er anders niet genoeg tijd is om weer te dalen en de volgende stijging vanaf een laag toonniveau te kunnen beginnen. De eerste stijging is dan vaak kleiner en vlakker dan de tweede.

In hoofdstuk vijf gaat het om twee vragen: Is de invloed, die naburige tonen op elkaar hebben, anders wanneer twee lettergrepen samen een zinsdeel (frase) vormen dan wanneer zij aparte frasen zijn? En zo ja, wat voor invloed heeft focus op dit verschil? Met betrekking tot de eerste vraag is het resultaat gecompliceerd: Er zijn wel verschillen in de mate van tooninvloed tussen twee lettergrepen binnen een frase vs. in twee aangrenzende frasen. Maar deze verschillen zijn meer zichtbaar als het gaat om invloed van de volgende op de voorafgaande toon, en er is minder invloed van de voorafgaande op de volgende toon. De soort constructie (een of twee frasen) en de positie van het woord waarnaar gekeken wordt (eerste of laatste positie) zijn dus beide belangrijk. Dit samenspel kan verklaard worden met hulp van prosodische theorie: Elk frase heeft maar een prosodisch “hoofd”, en als de twee lettergrepen samen een frase vormen, is alleen de laatste lettergreep een prosodisch “hoofd”. Prosodische “hoofden” worden minder beïnvloedt door aangrenzende tonen. De conclusie volgt dat prosodische “hoofdigheid” de belangrijkste factor is om tooncoarticulatie te verklaren.

Het tweede experiment in hoofdstuk vijf gaat opnieuw om de vraag hoe focus de gevonden toonverandering beïnvloedt. De resultaten laten zien dat de stimuli van experiment 1, als ze onder focus worden uitgesproken, allemaal minder door de context beïnvloed zijn. Toch is er nog steeds een verschil tussen tonen op lettergrepen in de eerste en in de laatste positie. Dit leidt tot de conclusie dat de invloed van focus op tooncoarticulatie verschilt van de invloed van prosodische structuur, omdat de laatste ook onder focus nog steeds te zien is. De twee effecten zijn dus additief, en niet varianten van hetzelfde effect. Andermaal is bevestigd dat focus de fonetische details (implementatie) van tooncontouren beïnvloedt zonder de fonologische categorie van de tonen zelf te veranderen.

Hoofdstuk zes gaat een stap verder en kijkt naar de implementatie van tooncontouren op het zinsniveau. Er bestaat onenigheid tussen onderzoekers die beweren dat sprekers hun hele zinnen vooraf plannen en het eerste accent in een zin hoger maken naar mate er een langere zin achteraan komt, en onderzoekers die geloven dat sprekers van accent tot accent plannen. Om de planning van zinnen in Wenzhou Chinees te onderzoeken, werd de lengte van verschillende zinsdelen zo gevariëerd dat precies kan worden nagegaan waar elk lengteverschil invloed heeft. Alle zinnen bestonden uit woorden met een stijgend-dalende tooncontour. Gemeten is hoe hoog de sprekers het eerste accent in elk zinsdeel uitspraken. De resultaten laten zien dat sprekers inderdaad de hoogte van de eerste accenten in elk zinsdeel anders kozen, maar uitsluitend afhankelijk van hoeveel woorden er in het testzinsdeel nog kwamen. De lengte van andere zinsdelen had geen enkele invloed op de accenthoogte in het testzinsdeel. Voor Wenzhou kan daarom worden gezegd dat de planning op niveau van zinsdelen plaats vindt en niet op het niveau van de hele zin.

Een tweede onderzoeksvraag in hoofdstuk zes is hoe complexe syntactische structuren in de toonrealisatie worden weergegeven. Zo hebben de zinnen *Mijn neef probeert de grammatica te oefenen* en *Mijn neef belooft de grammatica te oefenen* op het eerste gezicht dezelfde structuur. Maar de tweede zin kan worden uitgebreid tot *Mijn neef belooft zijn moeder de grammatica te oefenen*; met de eerste zin kan dit niet: **Mijn neef probeert zijn moeder de grammatica te oefenen*. Dit laat zien dat de twee zinnen verschillen in structuur, ook al is hun woordvolgorde aan de oppervlakte hetzelfde. De zin *Mijn neef belooft de grammatica te oefenen* bevat een verzwegen element (PRO) op de plek waar *zijn moeder* kan staan, waardoor deze structuur complexer is dan in het geval van *probeert*. In hoofdstuk zes is onderzocht of zulke

structuurverschillen weerspiegeld worden in de fonetische details (implementatie) van de tooncontouren.

Hoofdstuk zes laat inderdaad zien dat de implementatie van tooncontouren in complexe zinnen (zoals met *beloven)* verschilt van de uitvoering in minder complexe zinnen (zoals met *proberen*). Men was al eerder tot een vergelijkbare constatering gekomen voor zinnen in het Duits en in het Engels, maar het is interessant om te zien dat hetzelfde principe ook in het Wenzhou Chinees werkt.

Het laatste hoofdstuk voor de conclusie gaat nogmaals over manieren waarop sprekers iets benadrukken voor luisteraars en over welke categorieën hiervoor belangrijk zijn. In een eerste experiment wordt gekeken hoe sprekers plaatjes beschrijven als die plaatjes met verschillende voorafgaande vragen worden gepresenteerd. Zoals al eerder voor andere talen is gevonden, gebruiken de sprekers van het Wenzhou minder bepaalde en langere constructies om naar referenten te verwijzen die nieuw in de context zijn, en meer bepaalde en kortere constructies voor referenten die al uit de voorgaande context bekend zijn.

In een tweede experiment is gekeken hoe sprekers precies deze referenten markeren als de woordkeus vastligt op grond van een script. Het resultaat is dat de sprekers een drieledig verschil maken tussen referenten die bekend zijn, referenten die onderdeel zijn van een zin die helemaal nieuw is, en referenten die als zinsdeel specifiek benadrukt worden. Dit drieledig verschil kan worden teruggezien zowel in de hoogte van de tooncontour als in de lengte van de woorden waarmee naar de referenten verwezen wordt. Sprekers van het Wenzhou gebruiken dus de manier waarop ze tooncontouren realiseren om de luisteraar nauwkeurig informatie te geven over de bekendheid van een referent in de discours.

Hoofdstuk acht vat de resultaten van de experimenten samen en blikt vooruit op mogelijkheden voor toekomstig onderzoek.

Appendices

Appendix 2.1

Comparison between published tone sandhi descriptions and the current findings. Bold = (reasonable) match with contours in the previously published literature.

Citation forms	This thesis	Chen 2000	Hou 1998	Qian 1992	Rose 2000, 2002, 2004	You & Yang 1998
33+33	42.31	M.M	33.33	44.44	32.33	11.33
33+31	42.31	L.L	22.12	44.24	21.11	11.13
33+35	42.31	HM.MH	42.35	52.44	53.23	53.35
33+24	44.22	HM.MH	42.35	52.44	53.23	53.35
33+42	22.44	MLM.HM	22.42	44.52	**22.4**	13.53
33+11	22.35	MLM.HM	22.42	44.52	22.4	13.53
33+313	35.42	HM.Lq	35.13	25.24	**45.311**	53.13
33+212	35.42	HM.Lq	35.13	25.24	**45.311**	53.13
31+33	31.22	L.M	22.33	22.44	32.33	11.33
31+31	31.22	L.L	22.12	22.24	**21.11**	11.13
31+35	35.22	HM.MH	42.35	52.34	**343.23**	53.35
31+24	35.22	HM.MH	42.35	52.34	**343.23**	53.35
31+42	22.44	MLM.HM	22.42	22.52	**11.4**	13.53
31+11	22.35	MLM.HM	22.42	22.52	**11.34**	13.53
31+313	35.42	HM.Lq	-	25.24	**35.311**	53.13
31+212	35.42	HM.Lq	-	25.24	**35.311**	53.13
35+33	44.22	HM.M	42.33	52.44	53.22	42.33
35+31	42.31	**HM.ML**	**42.21**	**52.21**	**52.21**	**42.11**
35+35	44.22	HM.MH	42.35	52.34	53.23	53.35
35+24	44.22	HM.MH	42.35	52.34	53.23	53.35
35+42	42.31	**HM.ML**	**42.21**	**52.21**	**52.1**	**42.11**
35+11	44.22	HM.L	42.22	52.22	53.22	42.11
35+313	35.42	HM.Lq	35.13	25.24	**45.311**	53.13
35+212	35.42	HM.Lq	35.13	25.24	**45.311**	53.13
24+33	31.22	HM.M	42.33	52.44	**343.22**	42.33
24+31	31.22	HM.ML	**42.21**	52.21	**343.21**	42.11
24+35	35.22	HM.MH	42.35	52.34	343.23	53.35
24+24	35.22	HM.MH	42.35	52.34	343.23	53.35

24+42	31.22	HM.ML	**42.21**	52.21	**232.1**	42.11
24+11	31.22	HM.L	42.22	52.22	**343.22**	42.11
24+313	35.42	HM.Lq	35.13	25.24	**35.311**	53.13
24+212	35.42	HM.Lq	35.13	25.24	**35.311**	53.13
42+33	44.22	HM.M	42.33	52.44	53.22	42.33
42+31	42.31	L.L	22.12	44.24	32.12	11.13
42+35	44.22	HM.MH	-	52.34	53.23	53.35
42+24	44.22	HM.MH	-	52.34	53.23	53.35
42+42	42.31	**HM.ML**	**42.21**	**52.21**	**52.1**	**42.11**
42+11	44.22	HM.L	42.22	52.22	53.22	42.11
42+313	35.42	HM.Lq	35.13	25.24	**45.311**	53.13
42+212	35.42	HM.Lq	35.13	25.24	**45.311**	53.13
11+33	35.22	HM.M	42.33	52.44	343.22	42.33
11+31	31.22	HM.ML	-	52.21	**343.21**	42.11
11+35	35.22	HM.MH	42.35	52.34	343.23	53.35
11+24	35.22	HM.MH	42.35	52.34	343.23	53.35
11+42	31.22	HM.ML	-	52.21	**232.1**	42.11
11+11	35.22	HM.L	42.22	52.22	343.22	42.11
11+313	35.42	HM.Lq	35.13	25.24	**35.311**	53.13
11+212	35.42	HM.Lq	35.13	25.24	**35.311**	53.13
313+33	22.33	**Lq.M**	**21.33**	**3.44**	**1.33**	**1.33**
313+31	42.31	L.L	22.12	4.24	31.11	11.13
313+35	22.35	**Lq.MH**	**21.35**	**3.34**	**1.34**	**1.35**
313+24	22.35	**Lq.MH**	**21.35**	**3.34**	**1.114**	**1.35**
313+42	22.42	**Lq.HM**	**21.42**	**3.52**	**3.52**	**1.42**
313+11	42.31	Lq.L	21.22	3.22	3.33	1.11
313+313	35.42	Lq.Lq	21.13	3.24	4.311	1.13
313+212	35.42	Lq.Lq	21.13	3.24	4.311	1.13
212+33	22.33	**Lq.M**	**21.33**	**2.44**	**1.33**	**1.33**
212+31	42.31	L.L	22.12	2.24	21.11	11.13
212+35	22.35	**Lq.MH**	**21.35**	**2.34**	**1.34**	**1.35**
212+24	22.35	**Lq.MH**	**21.35**	**2.34**	**1.114**	**1.35**
212+42	22.42	**Lq.HM**	**21.42**	**2.52**	**2.52**	**1.42**
212+11	42.31	Lq.L	21.22	3.22	12.33	1.11
212+313	35.42	Lq.Lq	21.13	3.24	**23.211**	1.13
212+212	35.42	Lq.Lq	21.13	3.24	**23.211**	1.13

Citation forms	This thesis	You 2002	Zhang 2007a	Zhengzhang 1964a, b	Zhengzhang 2008
33+33	42.31	11.33	55.55	22.33	11.33
33+31	42.31	22.21	22.13	22.2	11.12
33+35	42.31	53.35	42.35	43.34/22.33	43.34
33+24	44.22	53.35	42.35	43.34	43.34
33+42	22.44	13.53	22.42	213.43	11.53
33+11	22.35	13.53	22.42	213.43	11.53
33+313	35.42	53.13	35.213	43.12	34.13
33+212	35.42	53.13	35.213	43.12	34.13
31+33	31.22	11.33	33.55	22.33	11.33
31+31	31.22	22.21	22.13	22.2/42.21	11.12
31+35	35.22	53.35	42.35	43.34	43.34
31+24	35.22	53.35	42.35	43.34	43.34
31+42	22.44	13.53	22.42	213.43	11.53
31+11	22.35	13.53	22.42	213.43	11.53
31+313	35.42	53.13	35.213	43.12	34.13
31+212	35.42	53.13	35.213	43.12	34.13
35+33	44.22	42.33	42.33	42.33	42.33
35+31	42.31	**42.21**	**42.21**	**42.21**/22.2	**42.21**
35+35	44.22	53.35	42.35	43.34	43.34
35+24	44.22	53.35	42.35	43.34	43.34
35+42	42.31	**42.21**	**42.21**	**42.21**	**42.21**
35+11	44.22	42.11	**44.22**	42.22	43.11
35+313	35.42	53.13	35.213	43.12	34.13
35+212	35.42	53.13	35.213	43.12	34.13
24+33	31.22	42.33	42.33	42.33	42.33
24+31	31.22	42.21	42.21	42.21/22.2	42.21
24+35	35.22	53.35	42.35	43.34	43.34
24+24	35.22	53.35	42.35	43.34	43.34
24+42	31.22	42.21	42.21	42.21	42.21
24+11	31.22	42.11	44.22	42.22	43.11
24+313	35.42	53.13	35.213	43.12	34.13
24+212	35.42	53.13	35.213	43.12	34.13
42+33	44.22	42.33	42.33	42.33	42.33

42+31	42.31	22.21	22.13	22.2/**42.21**	11.12
42+35	44.22	53.35	42.35	43.34	43.34
42+24	44.22	53.35	42.35	43.34	43.34
42+42	42.31	**42.21**	**42.21**	**42.21**	**42.21**
42+11	44.22	42.11	**44.22**	42.22	43.11
42+313	35.42	53.13	35.213	43.12	34.13
42+212	35.42	53.13	35.213	43.12	34.13
11+33	35.22	42.33	42.33	42.33/22.33	42.33
11+31	31.22	42.21	42.21	42.21/22.2	42.21
11+35	35.22	42.11	42.35	43.34	43.34
11+24	35.22	53.35	42.35	43.34	43.34
11+42	31.22	53.35	42.21	42.21	42.21
11+11	35.22	42.33	44.22	42.22/213.43	43.11
11+313	35.42	42.21	35.213	43.12	34.13
11+212	35.42	53.13	35.213	43.12	34.13
313+33	22.33	**1.33**	**22.55**	**21.33**	**01.33**
313+31	42.31	22.21	22.13	22.2	11.12
313+35	22.35	**1.35**	**22.35**	**21.34**	**01.34**
313+24	22.35	**1.35**	**22.35**	**21.34**	**01.34**
313+42	22.42	**1.42**	**22.42**	**21.42**	**01.42**
313+11	42.31	1.11	22.22	21.22	01.11
313+313	35.42	1.13	22.213	21.12	01.212
313+212	35.42	1.13	22.213	21.12	01.212
212+33	22.33	**1.33**	**22.55**	**21.33**	**01.33**
212+31	42.31	22.21	22.13	22.2	11.12
212+35	22.35	**1.35**	**22.35**	**21.34**	**01.34**
212+24	22.35	**1.35**	**22.35**	**21.34**	**01.34**
212+42	22.42	**1.42**	**22.42**	**21.42**/42.21	**01.42**
212+11	42.31	1.11	22.22	21.22	01.11
212+313	35.42	1.13	22.213	21.12	01.212
212+212	35.42	1.13	22.213	21.12	01.212

Appendix 2.2

Possible syllables in Wenzhou, after Zhengzhang 2008: 177ff. Dark shading: Unattested syllables. Light shading: Mismatch Zhengzhang 2008 – this thesis.

	a	**ɜ**	**ɵ**	**o**	**u**	**ø**	**i**	**ai**	**ɛi**	**au**	**ou**	**iɛ**	**uɔ**	**aŋ**	**ɜŋ**	**oŋ**
pʰ	pʰa	pʰɜ	pʰɵ	pʰo			pʰi	pʰai	pʰɛi			pʰiɛ	pʰuɔ	pʰaŋ	pʰɜŋ	pʰoŋ
p	pa	pɜ	pɵ	po			pi	pai	pɛi			piɛ	puɔ	paŋ	pɜŋ	poŋ
b	ba	bɜ	bɵ	bo			bi	bai	bɛi			biɛ	buɔ	baŋ	bɜŋ	boŋ
m	ma	mɜ	mɵ	mo			mi	mai	mɛi			miɛ	muɔ	maŋ	mɜŋ	moŋ
f	fa		fɵ	fo	fu			fai	fɛi		fou	fiɛ		faŋ		
v	va		vɵ	vo	vu			vai	vɛi			viɛ		vaŋ		
tʰ	tʰa	tʰɜ	tʰɵ				tʰi	tʰai	tʰɛi	tʰau	tʰou	tʰiɛ	tʰuɔ	tʰaŋ	tʰɜŋ	tʰoŋ
t	ta	tɜ	tɵ				ti	tai	tɛi	tau	tou	tiɛ	tuɔ	taŋ	tɜŋ	toŋ
d	da	dɜ	dɵ				di	dai	dɛi	dau	dou	diɛ	duɔ	daŋ	dɜŋ	doŋ
tsʰ	tsʰa	tsʰɜ	tsʰɵ	tsʰo			tsʰi	tsʰai	tsʰɛi	tsʰau	tsʰou	tsʰiɛ	tsʰuɔ	tsʰaŋ	tsʰɜŋ	tsʰoŋ
ts	tsa	tsɜ	tsɵ	tso			tsi	tsai	tsɛi	tsau	tsou	tsiɛ	tsuɔ	tsaŋ	tsɜŋ	tsoŋ
dz	dza	dzɜ	dzɵ	dzo			dzi	dzai	dzɛi	dzau	dzou	dziɛ	dzuɔ	dzaŋ	dzɜŋ	dzoŋ
n	na	nɜ	nɵ	no		nø	ni	nai	nɛi	nau	nou	niɛ	nuɔ	naŋ		noŋ
s	sa	sɜ	sɵ	so			si	sai	sɛi	sau	sou	siɛ	suɔ	saŋ	sɜŋ	soŋ
z	za	zɜ	zɵ	zo	zu		zi	zai	zɛi	zau	zou		zuɔ	zaŋ	zɜŋ	zoŋ
l	la	lɜ	lɵ	lo			li	lai	lɛi	lau	lou	liɛ	luɔ	laŋ	lɜŋ	loŋ
tɕʰ	tɕʰa			tɕʰo		tɕʰø	tɕʰi	tɕʰai		tɕʰau	tɕʰou	tɕʰiɛ	tɕʰuɔ	tɕʰaŋ		tɕʰoŋ
tɕ	tɕa			tɕo		tɕø	tɕi	tɕai		tɕau	tɕou	tɕiɛ	tɕuɔ	tɕaŋ		tɕoŋ
dʑ	dʑa			dʑo		dʑø	dʑi	dʑai		dʑau	dʑou	dʑiɛ	dʑuɔ	dʑaŋ		dʑoŋ
ɲ	ɲa			ɲo		ɲø	ɲi	ɲai		ɲau	ɲou	ɲiɛ	ɲuɔ	ɲaŋ		
ɕ	ɕa			ɕo		ɕø	ɕi	ɕai		ɕau	ɕou	ɕiɛ	ɕuɔ	ɕaŋ		ɕoŋ
j	ja			jo		jø	ji	jai		jau	jou	jiɛ	juɔ	jaŋ		joŋ
kʰ	kʰa	kʰɜ	kʰɵ	kʰo	kʰu	kʰø		kʰai	kʰɛi	kʰau		kʰiɛ	kʰuɔ	kʰaŋ		kʰoŋ
k	ka	kɜ	kɵ	ko	ku	kø	ki	kai	kɛi	kau		kiɛ	kuɔ	kaŋ		koŋ
g	ga	gɜ	gɵ	go	gu	gø	gi	gai	gɛi	gau		giɛ	guɔ	gaŋ		goŋ
ŋ	ŋa	ŋɜ	ŋɵ	ŋo	ŋu			ŋai	ŋɛi	ŋau		ŋiɛ	ŋuɔ			
ʔ	ʔa	ʔɜ	ʔɵ	ʔo	ʔu	ʔø	ʔi	ʔai	ʔɛi	ʔau			ʔuɔ	ʔaŋ		ʔoŋ
h	ha	hɜ	hɵ	ho			hi		hɛi	hau			huɔ	haŋ		hoŋ
ɦ	ɦa	ɦɜ	ɦɵ	ɦo			ɦi		ɦɛi	ɦau		ɦiɛ	ɦuɔ	ɦaŋ		ɦoŋ

Appendix 3.1

Analysis of degree of lexicalization of the stimuli used in Chapter 3 experiments 1 and 2.

Criteria:

- M = Morphology: One of the two morphemes is bound
- Se = Semantics: The meaning of the phrase is idiomatized and cannot (fully) be retrieved from the meaning of the two parts
- Sy = Syntax: The object cannot be freely moved within the sentence, for example in a preverbal position
- Lex = Degree of lexicalization, expressed numerically

Experiment 1

Hanzi	Citation forms		Translation	M	Se	Sy	Lex
买车	*ma24*	*ts^{h}u33*	'to buy a car'	/	/	/	0
卖车	*ma11*	*ts^{h}u33*	'to sell a car'	/	/	/	0
卖菜	*ma11*	*ts^{h}ɛi42*	'to sell vegetables'	/	/	/	0
写信	*sɛi35*	*saŋ42*	'to write a letter'	/	/	/	0
踢球	*t^{h}ɛi42*	*dʑau31*	'to kick a ball'	/	/	/	0
找人	*zaŋ11*	*naŋ31*	'to look for a person'	/	/	/	0
喝水	*ha42*	*sʐ̩35*	'to drink water'	/	/	/	0
喝酒	*ha42*	*tɕou35*	'to drink alcohol'	/	/	/	0
退票	*t^{h}ai42*	*p^{h}iɛ42*	'to return a ticket'	/	/	/	0
写字	*sɛi35*	*zʐ̩11*	'to write characters'	/	/	/	0
唱戏	*tɕ^{h}i42*	*sʐ̩42*	'to sing opera'	/	/	/	0
打人	*tiɛ35*	*naŋ31*	'to hit a person'	/	/	/	0
开门	*k^{h}i33*	*maŋ31*	'to open a door'	/	/	/	0
煮饭	*tsɛi35*	*va11*	'to cook rice'	/	/	/	0
煮菜	*tsɛi35*	*ts^{h}ɛi42*	'to prepare food'	/	/	/	0
煮面	*tsɛi35*	*mi11*	'to cook noodles'	/	/	/	0
停车	*deŋ31*	*ts^{h}u33*	'to park a car'	/	√	/	1

下雨	*ɦu24*	*vu24*	'to rain'	/	√	/	1
看书	*tʰsʐ42*	*sʐ33*	'to read a book'	/	√	/	1
唱歌	*tɕʰi42*	*ku33*	'to sing a song'	/	√	/	1
说话	*kuɔ33*	*ɦu11*	'to speak'	/	√	/	1
开车	*kʰi33*	*tsʰu33*	'to drive a car'	/	√	/	1
打球	*tiɛ35*	*dʑau31*	'to play a ballgame'	/	√	/	1
开灯	*kʰi33*	*taŋ33*	'to switch on a lamp'	/	√	/	1
做饭	*tsou42*	*va11*	'to prepare a meal'	/	√	/	1
跳舞	*tʰiɛ42*	*vu24*	'to dance'	/	√	/	1
等车	*taŋ35*	*tsʰu33*	'to wait for the bus'	/	/	√	1
回家	*vai31*	*ku33*	'to return home'	/	/	√	1
回国	*vai31*	*kai313*	'to return home'	/	/	√	1
洗澡	*sɛi35*	*ju212*	'to take a bath'	√	√	/	2
考试	*kʰɜ35*	*sʐ42*	'to take an exam'	√	√	/	2
闭门	*pɛi42*	*maŋ31*	'to shut the door'	√	/	√	2
跑步	*pʰuɔ35*	*bɵ11*	'to run'	/	√	√	2
出国	*tɕʰø313*	*kai313*	'to go abroad'	/	√	√	2
出门	*tɕʰø313*	*maŋ31*	'to go out'	/	√	√	2
坐车	*zu24*	*tsʰu33*	'to take a bus'	/	√	√	2
教书	*kuɔ42*	*sʐ33*	'to teach'	/	√	√	2
走路	*tsau24*	*lɵ11*	'to walk'	/	√	√	2
念书	*ni11*	*sʐ33*	'to study a book'	/	√	√	2
升旗	*sɜŋ33*	*dzʐ31*	'to raise a flag'	/	√	√	2
跑马	*pʰuɔ35*	*mu24*	'to ride a horse'	/	√	√	2
摇头	*ɦiɛ11*	*dou31*	'to shake one's head'	/	√	√	2
斗鸡	*tau42*	*tsʐ33*	'cockfighting'	/	√	√	2
见面	*tɕi42*	*mi11*	'to meet'	/	√	√	2
变心	*pi42*	*saŋ33*	'to be unfaithful'	/	√	√	2

Experiment 2

Hanzi	Citation forms		Translation	M	Se	Sy	Lex
开灯	*kʰi33*	*taŋ33*	'to switch on a lamp'	/	/	/	0
煮肉	*tsɛi35*	*ɻ̩ou212*	'to cook meat'	/	/	/	0
包书	*puɔ33*	*sz̩33*	'to wrap a book'	/	/	/	0
包机	*puɔ33*	*tsz̩33*	'to charter a plane'	/	√	/	1
救国	*tɕau42*	*kai313*	'to save the nation'	/	/	√	1
建国	*tɕi42*	*kai313*	'to found a state'	/	/	√	1
卖国	*ma11*	*kai313*	'to betray one's country'	/	√	√	2
收心	*ɕou33*	*saŋ33*	'to collect one's thoughts'	/	√	√	2

Appendix 3.2

Stimuli from experiment 2 in Chapter 3. Tone numbers in the Wenzhou transcriptions refer to the citation tones on the syllables.

Hanzi	Citation forms		Translation	Type
分队	*faŋ33*	*dai11*	'unit, gang'	disyllabic noun
分内	*faŋ33*	*nai11*	'duty'	disyllabic noun
资本	*tsʐ33*	*paŋ35*	'capital'	disyllabic noun
书本	*sʐ33*	*paŋ35*	'books'	disyllabic noun
科长	*ku33*	*tɕi35*	'section chief'	disyllabic noun
家长	*ku33*	*tɕi35*	'head of a household'	disyllabic noun
鸡心	*tsʐ33*	*saŋ33*	'chicken heart'	disyllabic noun
机心	*tsʐ33*	*saŋ33*	'watch movement'	disyllabic noun
公园	*koŋ33*	*jø31*	'park, garden'	disyllabic noun
公元	*koŋ33*	*nø31*	'Christian era; A.D.'	disyllabic noun
石头	*zɛi212*	*dou31*	'stone'	disyllabic noun
食堂	*zɛi212*	*duɔ31*	'canteen'	disyllabic noun
鱼肉	*ŋɵ212*	*ɲou212*	'fish and meat'	disyllabic noun
骨肉	*kɵ313*	*ɲou212*	'flesh and blood'	disyllabic noun
歪心	*va33*	*saŋ33*	'twisted mind'	disyllabic noun
外心	*va11*	*saŋ33*	'marital infidelity'	disyllabic noun
包抄	*puɔ33*	*tsʰuɔ33*	'to outflank'	disyllabic verb
包车	*puɔ33*	*tsʰu33*	'chartered vehicle'	disyllabic noun
分布	*faŋ33*	*pɵ42*	'to be scattered'	disyllabic verb
分寸	*faŋ33*	*tsʰɵ42*	'proper restraint'	disyllabic noun
大方	*dou11*	*huɔ33*	'generous'	disyllabic adjective
地方	*dɛi11*	*huɔ33*	'place, room'	disyllabic noun
开始	*kʰi33*	*sʐ35*	'to begin; start'	disyllabic verb
开水	*kʰi33*	*sʐ35*	'boiled water'	disyllabic noun
公报	*koŋ33*	*pɔ42*	'bulletin; gazette'	disyllabic noun
公布	*koŋ33*	*pɵ42*	'to announce'	disyllabic verb
关心	*ka33*	*saŋ33*	'to be concerned'	disyllabic verb
丹心	*ta33*	*saŋ33*	'loyalty'	disyllabic noun

外国	*va11*	*kai313*	‘foreign country’	disyllabic noun
卖国	*ma11*	*kai313*	‘to betray one's country’	VO-construction
战国	*tɕi42*	*kai313*	‘Warring States period’	disyllabic noun
救国	*tɕau42*	*kai313*	‘to save the nation’	VO-construction
建国	*tɕi42*	*kai313*	‘to found a state’	VO-construction
鼠肉	*ts^{h}ɔi35*	*ʃou212*	‘rat meat’	disyllabic noun
煮肉	*tsɔi35*	*ʃou212*	‘to cook meat’	VO-construction
手心	*ɕou35*	*saŋ33*	‘palm of the hand’	disyllabic noun
收心	*ɕou33*	*saŋ33*	‘to collect thoughts’	VO-construction
决心	*tɕø313*	*saŋ33*	‘determination’	disyllabic noun
开灯	*k^{h}i33*	*taŋ33*	‘to turn on the light’	VO-construction
包机	*puɔ33*	*tsʐ̩33*	‘charter a plane’	VO-construction
包书	*puɔ33*	*sʐ̩33*	‘wrap a book’	VO-construction

Appendix 4.1

List of stimuli for the experiment reported in Chapter 4. "Target word" lists the citation forms of the syllables that form the tone sandhi domain. Bolding indicates contrast location in the context questions.

HH register

a. Target word: 中国 *tɕoŋ33 kai313* 'China'

Target sentence:	不，	我说	中国	大学	这几个字眼.
	fu	*ŋ kuɔ*	*tɕoŋ kai*	*da ɦu*	*ki ki kai zɻ ŋa*
	no	I say	TARGET	FRAME	these words

'No, I say the words Chinese university.'

First syllable focus:	你	说	**美**国	大学	啊？
	ni	*kuɔ*	***mɛi*** *kai*	*da ɦu*	*a*
	you	say	TARGET	FRAME	Q

'Are you saying **American** university?'

Second syllable focus:	你	说	中**央**	大学	啊？
	ni	*kuɔ*	*tɕoŋ* ***ʔi***	*da ɦu*	*a*
	you	say	TARGET	FRAME	Q

'Are you saying **central** university?'

Whole target focus:	你	说	**外语**	大学	啊？
	ni	*kuɔ*	***va nø***	*da ɦu*	*a*
	you	say	TARGET	FRAME	Q

'Are you saying **foreign language** university?'

Pre-target focus:	你	**写**	中国	大学	啊?
	ni	***ɕi***	*tɕoŋ kai*	*da ɦu*	*a*
	you	write	TARGET	FRAME	Q

'Are you **writing** Chinese university?'

Post-target focus:	你	说	中国	**政府**	啊？
	ni	*kuɔ*	*tɕoŋ kai*	***dzɜŋ fø***	*a*
	you	say	TARGET	FRAME	Q

'Are you saying Chinese **government**?'

b. Target word: 战国 *tɕi42 kai313* ‘Warring states period’

Target sentence:	不，	我说	战国	小说	这几个字眼.
	fu	*ŋ kuɔ*	*tɕi kai*	*ɕiɛ ɕø*	*ki ki kai zɻ ŋa*
	no	I say	TARGET	FRAME	these words

‘No, I say the words Warring states period novel.’

First syllable focus:	你	说	**法**国	小说	啊？
	ni	*kuɔ*	***ho*** *kai*	*ɕiɛ ɕø*	*a*
	you	say	TARGET	FRAME	Q

‘Are you saying **French** novel?’

Second syllable focus:	你	说	战**争**	小说	啊？
	ni	*kuɔ*	*tɕi* ***tsiɛ***	*ɕiɛ ɕø*	*a*
	you	say	TARGET	FRAME	Q

‘Are you saying **War** novel?’

Whole target focus:	你	说	**言情**	小说	啊？
	ni	*kuɔ*	***ni zeŋ***	*ɕiɛ ɕø*	*a*
	you	say	TARGET	FRAME	Q

‘Are you saying **romantic** novel?’

Pre-target focus:	你	**写**	战国	小说	啊？
	ni	***ɕi***	*tɕi kai*	*ɕiɛ ɕø*	*a*
	you	write	TARGET	FRAME	Q

‘Are you **writing** Warring states period novel?’

Post-target focus:	你	说	战国	**戏剧**	啊？
	ni	*kuɔ*	*tɕi kai*	***sɜ tɕai***	*a*
	you	say	TARGET	FRAME	Q

‘Are you saying Warring states period **play**?’

c. Target word: 教室 *kuɔ42 sai313* ‘classroom’

Target sentence:	不，	我说	教室	地板	这几个字眼.
	fu	*ŋ kuɔ*	*kuɔ sai*	*dɛi pa*	*ki ki kai zɻ ŋa*
	no	I say	TARGET	FRAME	these words

‘No, I say the words classroom floorboard.’

First syllable focus:	你	说	**卧**室	地板	啊？
	ni	*kuɔ*	***ŋ*** *sai*	*dɛi pa*	*a*
	you	say	TARGET	FRAME	Q
	‘Are you saying **bed**room floorboard?’				
Second syllable focus:	你	说	教**堂**	地板	啊？
	ni	*kuɔ*	*kuɔ* ***duɔ***	*dɛi pa*	*a*
	you	say	TARGET	FRAME	Q
	‘Are you saying **church** floorboard?’				
Whole target focus:	你	说	**厕所**	地板	啊？
	ni	*kuɔ*	***tsɛi su***	*dɛi pa*	*a*
	you	say	TARGET	FRAME	Q
	‘Are you saying **bathroom** floorboard?’				
Pre-target focus:	你	**写**	教室	地板	啊？
	ni	***ɕi***	*kuɔ sai*	*dɛi pa*	*a*
	you	write	TARGET	FRAME	Q
	‘Are you **writing** classroom floorboard?’				
Post-target focus:	你	说	教室	**窗户**	啊？
	ni	*kuɔ*	*kuɔ sai*	***tɕʰuɔ ma***	*a*
	you	say	TARGET	FRAME	Q
	‘Are you saying classroom **window**?’				

HL register

d. Target word: 短袜 *tɵ35 mu212* ‘socks’

Target sentence: 不，	我说	短袜	商店	这几个字眼.	
fu	*ŋ kuɔ*	*tɵ mu*	*ɕi ti*	*ki ki kai zɻ̩ ŋa*	
no	I say	TARGET	FRAME	these words	
‘No, I say the words sock shop.’					

First syllable focus:	你	说	**毛**袜	商店	啊？
	ni	*kuɔ*	***mɜ*** *mu*	*ɕi ti*	*a*
	you	say	TARGET	FRAME	Q
	‘Are you saying **stocking** shop?’				

Second syllable focus:	你	说	短**袄**	商店	啊？
	ni	*kuɔ*	*tɵ ʔɜ*	*ʨi ti*	*a*
	you	say	TARGET	FRAME	Q
	‘Are you saying **short jacket** shop?’				
Whole target focus:	你	说	**内衣**	商店	啊？
	ni	*kuɔ*	***nai ji***	*ʨi ti*	*a*
	you	say	TARGET	FRAME	Q
	‘Are you saying **underwear** shop?’				
Pre-target focus:	你	**写**	短袜	商店	啊？
	ni	***ʨi***	*tɵ mu*	*ʨi ti*	*a*
	you	write	TARGET	FRAME	Q
	‘Are you **writing** sock shop?’				
Post-target focus:	你	说	短袜	**公司**	啊？
	ni	*kuɔ*	*tɵ mu*	***koŋ sʐ̩***	*a*
	you	say	TARGET	FRAME	Q
	‘Are you saying sock **company**?’				

e. Target word: 中学 *ʨoŋ33 ɦu212* ‘Middle school’

Target sentence: 不，	我说	中学		考试	这几个字眼.
fu	*ŋ kuɔ*	*ʨoŋ ɦu*		*kɜ sʐ̩*	*ki ki kai zʐ̩ ŋa*
no	I say	TARGET		FRAME	these words
‘No, I say the words Middle school exam.’					

First syllable focus:	你	说	**大**学	考试	啊？
	ni	*kuɔ*	***da*** *ɦu*	*kɜ sʐ̩*	*a*
	you	say	TARGET	FRAME	Q
	‘Are you saying **university** exam?’				
Second syllable focus:	你	说	中**文**	考试	啊？
	ni	*kuɔ*	*ʨoŋ* ***vaŋ***	*kɜ sʐ̩*	*a*
	you	say	TARGET	FRAME	Q
	‘Are you saying **Chinese** exam?’				
Whole target focus:	你	说	**外语**	考试	啊？
	ni	*kuɔ*	***va nø***	*kɜ sʐ̩*	*a*
	you	say	TARGET	FRAME	Q
	‘Are you saying **foreign language** exam?’				

Pre-target focus:	你	**写**	中学	考试	啊？
	ni	***ɕi***	*tɕoŋ ʃu*	*kɔ sʐ*	*a*
	you	write	TARGET	FRAME	Q
	‘Are you **writing** Middle school exam?’				
Post-target focus:	你	说	中学	**校长**	啊？
	ni	*kuɔ*	*tɕoŋ ʃu*	***vuɔ tɕi***	*a*
	you	say	TARGET	FRAME	Q
	‘Are you saying Middle school **principal**?’				

f. Target word: 小麦 *ɕiɛ35 ma212* ‘wheat’

Target sentence:	不，	我说	小麦	商店	这几个字眼.
	fu	*ŋ kuɔ*	*ɕiɛ ma*	*ɕi ti*	*ki ki kai zʐ ŋa*
	no	I say	TARGET	FRAME	these words
	‘No, I say the words wheat shop.’				

First syllable focus:	你	说	**大**麦	商店	啊？
	ni	*kuɔ*	***da*** *ma*	*ɕi ti*	*a*
	you	say	TARGET	FRAME	Q
	‘Are you saying **barley** shop?’				
Second syllable focus:	你	说	小**米**	商店	啊？
	ni	*kuɔ*	*ɕiɛ mɛi*	*ɕi ti*	*a*
	you	say	TARGET	FRAME	Q
	‘Are you saying **millet** shop?’				
Whole target focus:	你	说	**蔬菜**	商店	啊？
	ni	*kuɔ*	***sou tsɛi***	*ɕi ti*	*a*
	you	say	TARGET	FRAME	Q
	‘Are you saying **vegetable** shop?’				
Pre-target focus:	你	**写**	小麦	商店	啊？
	ni	***ɕi***	*ɕiɛ ma*	*ɕi ti*	*a*
	you	write	TARGET	FRAME	Q
	‘Are you **writing** wheat shop?’				
Post-target focus:	你	说	小麦	**工厂**	啊？
	ni	*kuɔ*	*ɕiɛ ma*	***koŋ ts^{h}ɛi***	*a*
	you	say	TARGET	FRAME	Q
	‘Are you saying wheat **factory**?’				

LH register

g. Target word: 牙刷 *ŋu31 sɵ313* 'toothbrush'

Target sentence: 不， 我说 牙刷 商店 这几个字眼.
fu *ŋ kuɔ* *ŋu sɵ* *ɕi ti* *ki ki kai zɻ ŋa*
no I say TARGET FRAME these words
'No, I say the words toothbrush shop.'

First syllable focus: 你 说 **鞋**刷 商店 啊?
ni *kuɔ* ***ɦa*** *sɵ* *ɕi ti* *a*
you say TARGET FRAME Q
'Are you saying **shoe** brush shop?'

Second syllable focus: 你 说 牙**膏** 商店 啊?
ni *kuɔ* *ŋu* ***kɜ*** *ɕi ti* *a*
you say TARGET FRAME Q
'Are you saying tooth**paste** shop?'

Whole target focus: 你 说 **短袜** 商店 啊?
ni *kuɔ* *tɵ mu* *ɕi ti* *a*
you say TARGET FRAME Q
'Are you saying **sock** shop?'

Pre-target focus: 你 **写** 牙刷 商店 啊?
ni ***ɕi*** *ŋu sɵ* *ɕi ti* *a*
you write TARGET FRAME Q
'Are you **writing** toothbrush shop?'

Post-target focus: 你 说 牙刷 **容器** 啊?
ni *kuɔ* *ŋu sɵ* ***joŋ tsʰɻ*** *a*
you say TARGET FRAME Q
'Are you saying toothbrush **container**?'

h. Target word: 外国 *va11 kai313* 'foreign country'

Target sentence: 不， 我说 外国 大学 这几个字眼.
fu *ŋ kuɔ* *va kai* *da ɦu* *ki ki kai zɻ ŋa*
no I say TARGET FRAME these words
'No, I say the words foreign country university.'

First syllable focus:	你	说	**中**国	大学	啊？
	ni	*kuɔ*	***tɕoŋ kai***	*da ɦu*	*a*
	you	say	TARGET	FRAME	Q
	‘Are you saying **Chinese** university?’				
Second syllable focus:	你	说	外**语**	大学	啊？
	ni	*kuɔ*	*va* ***nø***	*da ɦu*	*a*
	you	say	TARGET	FRAME	Q
	‘Are you saying foreign **language** university?’				
Whole target focus:	你	说	**美术**	大学	啊？
	ni	*kuɔ*	***mɛi ju***	*da ɦu*	*a*
	you	say	TARGET	FRAME	Q
	‘Are you saying **arts** university?’				
Pre-target focus:	你	**写**	外国	大学	啊？
	ni	***ɕi***	*va kai*	*da ɦu*	*a*
	you	write	TARGET	FRAME	Q
	‘Are you writing foreign country university?’				
Post-target focus:	你	说	外国	**政府**	啊？
	ni	*kuɔ*	*va kai*	***dzɜŋ fø*** *a*	
	you	say	TARGET	FRAME	Q
	‘Are you saying foreign country **government**?’				

i. Target word: 蜡烛 *ɦɛi212 sɛi313* ‘candle’

Target sentence:	不，	我说	蜡烛	商店	这几个字眼.
	fu	*ŋ kuɔ*	*ɦɛi sɛi*	*ɕi ti*	*ki ki kai zɻ ŋa*
	no	I say	TARGET	FRAME	these words
	‘No, I say the words candle shop.’				

First syllable focus:	你	说	**灯**烛	商店	啊？
	ni	*kuɔ*	***taŋ*** *sɛi*	*ɕi ti*	*a*
	you	say	TARGET	FRAME	Q
	‘Are you saying **lamp** shop?’				
Second syllable focus:	你	说	蜡**纸**	商店	啊？
	ni	*kuɔ*	*ɦɛi* ***tsɛi***	*ɕi ti*	*a*
	you	say	TARGET	FRAME	Q
	‘Are you saying **wax paper** shop?’				

Whole target focus:	你	说	**短袜**	商店	啊？
	ni	*kuɔ*	***tɵ mu***	*ɕi ti*	*a*
	you	say	TARGET	FRAME	Q
	‘Are you saying **socks** shop?’				
Pre-target focus:	你	**写**	蜡烛	商店	啊？
	ni	***ɕi***	*ɦɛi sɛi*	*ɕi ti*	*a*
	you	write	TARGET	FRAME	Q
	‘Are you **writing** candle shop?’				
Post-target focus:	你	说	蜡烛	**容器**	啊？
	ni	*kuɔ*	*ɦɛi sɛi*	***joŋ tsʰɻ***	*a*
	you	say	TARGET	FRAME	Q
	‘Are you saying candle **container**?’				

j. Target word: 语法 *nø24 hu212* ‘grammar’

Target sentence: 不，	我说	语法	考试	这几个字眼.	
fu	*ŋ kuɔ*	*nø hu*	*kɜ sɻ*	*ki ki kai zɻ ŋa*	
no	I say	TARGET	FRAME	these words	
‘No, I say the words grammar exam.’					

First syllable focus:	你	说	**宪**法	考试	啊？
	ni	*kuɔ*	***ɕi*** *hu*	*kɜ sɻ*	*a*
	you	say	TARGET	FRAME	Q
	‘Are you saying **constitution** exam?’				
Second syllable focus:	你	说	语**言**	考试	啊？
	ni	*kuɔ*	*nø* ***ni***	*kɜ sɻ*	*a*
	you	say	TARGET	FRAME	Q
	‘Are you saying **language** exam?’				
Whole target focus:	你	说	**数学**	考试	啊？
	ni	*kuɔ*	*sɵ ɦu*	*kɜ sɻ*	*a*
	you	say	TARGET	FRAME	Q
	‘Are you saying **mathematics** exam?’				
Pre-target focus:	你	**写**	语法	考试	啊？
	ni	***ɕi***	*nø hu*	*kɜ sɻ*	*a*
	you	write	TARGET	FRAME	Q
	‘Are you **writing** grammar exam?’				

Post-target focus:	你	说	语法	**老师**	啊？
	ni	*kuɔ*	*nø hu*	***lɜ sz̩***	*a*
	you	say	TARGET	FRAME	Q

‘Are you saying grammar **teacher**?’

LL register

k. Target word: 文学 *vaŋ31 ɦu212* ‘literature’

Target sentence: 不，	我说	文学	考试	这几个字眼.	
fu	*ŋ kuɔ*	*vaŋ ɦu*	*kɜ sz̩*	*ki ki kai zz̩ ŋa*	
no	I say	TARGET	FRAME	these words	

‘No, I say the words literature exam.’

First syllable focus:	你	说	**大**学	考试	啊？
	ni	*kuɔ*	***da ɦu***	*kɜ sz̩*	*a*
	you	say	TARGET	FRAME	Q

‘Are you saying **university** exam?’

Second syllable focus:	你	说	文**法**	考试	啊？
	ni	*kuɔ*	*vaŋ* ***nø***	*kɜ sz̩*	*a*
	you	say	TARGET	FRAME	Q

‘Are you saying **grammar** exam?’

Whole target focus:	你	说	**外语**	考试	啊？
	ni	*kuɔ*	***va nø***	*kɜ sz̩*	*a*
	you	say	TARGET	FRAME	Q

‘Are you saying foreign language exam?’

Pre-target focus:	你	**写**	文学	考试	啊？
	ni	***ɕi***	*vaŋ ɦu*	*kɜ sz̩*	*a*
	you	write	TARGET	FRAME	Q

‘Are you **writing** literature exam?’

Post-target focus:	你	说	文学	**杂志**	啊？
	ni	*kuɔ*	*vaŋ ɦu*	***zø tsz̩***	*a*
	you	say	TARGET	FRAME	Q

‘Are you saying literature **magazine**?’

l. Target word: 腊肉 *la212 ŋou212* ‘bacon’

Target sentence:	不，	我说	腊肉	菜饭	这几个字眼.
	fu	*ŋ ku ɔ*	*la ɳou*	*ts^{h}ɛi va*	*ki ki kai zɻ ɳa*
	no	I say	TARGET	FRAME	these words

‘No, I say the words bacon dish.’

First syllable focus:	你	说	**牛**肉	菜饭	啊？
	ni	*ku ɔ*	***ŋau ɳou***	*ts^{h}ɛi va*	*a*
	you	say	TARGET	FRAME	Q

‘Are you saying **beef** dish?’

Second syllable focus:	你	说	腊**肠**	菜饭	啊？
	ni	*ku ɔ*	*la* ***dʑi***	*ts^{h}ɛi va*	*a*
	you	say	TARGET	FRAME	Q

‘Are you saying **sausage** dish?’

Whole target focus:	你	说	**豆腐**	菜饭	啊？
	ni	*ku ɔ*	***dou vø***	*ts^{h}ɛi va*	*a*
	you	say	TARGET	FRAME	Q

‘Are you saying **tofu** dish?’

Pre-target focus:	你	**写**	腊肉	菜饭	啊？
	ni	***ɕi***	*la ɳou*	*ts^{h}ɛi va*	*a*
	you	write	TARGET	FRAME	Q

‘Are you **writing** bacon dish?’

Post-target focus:	你	说	腊肉	**饭店**	啊？
	ni	*ku ɔ*	*la ɳou*	***va ti***	*a*
	you	say	TARGET	FRAME	Q

‘Are you saying bacon **restaurant**?’

Appendix 5.1

Experimental stimuli used in experiments 1 and 2 in Chapter 5. The following list depicts the experimental stimuli as used in both experiments. Information given (from left to right):

Hanzi: Chinese characters
MS: Morphosyntactic structure of the phrase
Citation forms: Broad transcription of target and context syllables
Target: Lexical tone on target word
Context: Tone sandhi contour on context word
1/0: Conflicting context (1) or compatible context (0)
Translation: English translation of phrase

For the stimuli with the target syllable in final position, the two columns giving the tone information are reversed to reflect the succession of tones in the stimulus phrase.

Hanzi	MS	Citation forms	Target	Context	1/0	Translation
已练习	AVV	*ji (li-zai)*	24	35-31	1	'practice already'
买皮夹	VNN	*ma (pɛi-ka)*	24	35-42	1	'buy a wallet'
已公布	AVV	*ji (koŋ-pɵ)*	24	22-33	1	'announce already'
买黄豆	VNN	*ma (ʎuɔ-tou)*	24	22-33	1	'buy soy beans'
总喜欢	AVV	*dzoŋ (sʐ-cø)*	24	42-31	0	'like always'
买汤圆	VNN	*ma (p^huɔ-jø)*	24	42-31	0	'buy dumpling soup'
已见面	AVV	*ji (tɕi-mi)*	24	44-22	0	'meet already'
动畚斗	VNN	*doŋ (paŋ-tau)*	24	44-22	0	'move the trashcan'
还奋斗	AVV	*va (faŋ-tau)*	31	42-22	1	'fight again'
回温洲	VNN	*vai (jø-tɕou)*	31	42-31	1	'return to Wenzhou'
还失败	AVV	*va (sai-ba)*	31	33-22	1	'lose again'
回南京	VNN	*vai (nɵ-tɕaŋ)*	31	33-22	1	'return to Nanjing'
还接收	AVV	*va (tɕi-ɕou)*	31	22-33	0	'receive again'
研黄豆	VNN	*ni (ʎuɔ-tou)*	31	22-33	0	'grind soy beans'
还重设	AVV	*va (dʑuɔ-sɛi)*	31	35-42	0	'rebuild again'
回中国	VNN	*vai (tɕoŋ-kai)*	31	35-42	0	'return to China'
好学习	AVV	*hɜ (ɦu-zai)*	35	35-31	1	'easy to learn'

想中国 VNN	*ɕi (tɕoŋ-kai)*	35	35-42	1	‘miss China’
好翻录 AVV	*hɜ (fa-lu)*	35	35-31	1	‘easy to duplicate’
打皇后 VNN	*tiɛ (va-ɦou)*	35	35-22	1	‘hit the empress’
弗广播 AVV	*fu (kuɔ-pɵ)*	35	44-22	0	‘not broadcast’
想广州 VNN	*ɕi (kuɔ-tɕou)*	35	44-22	0	‘miss Guangzhou’
好见面 AVV	*hɜ (tɕi-mi)*	35	44-22	0	‘meet easily’
打对手 VNN	*tiɛ (tai-tɕou)*	35	44-22	0	‘hit the opponent’
都喜欢 AVV	*ʔu (sʐ-ɕø)*	42	42-31	1	‘like everything’
喝汤圆 VNN	*ha (t^huɔ-jø)*	42	42-31	1	‘drink dumpling soup’
再见面 AVV	*tsɛi (tɕi-mi)*	42	44-22	1	‘meet again’
到广州 VNN	*tɜ (kuɔ-tɕou)*	42	44-22	1	‘reach Guangzhou’
再印刷 AVV	*tsɛi (jaŋ-sɵ)*	42	35-42	0	‘print again’
到中国 VNN	*tɜ (tɕoŋ-kai)*	42	35-42	0	‘reach China’
必学习 AVV	*pi (ɦu-zai)*	42	44-22	0	‘certainly learn’
剁猪肉 VNN	*tou (tsɛi-ɲou)*	42	44-22	0	‘chop pork’

Hanzi MS	Citation forms	Context	Target	1/0	Translation
干脆买 AAV	*(kø-ts^hai) ma*	22-33	24	1	‘simply buy’
接收米 VVN	*(tɕi-ɕou) mɛi*	22-33	24	1	‘receive rice’
当面买 AAV	*(tuɔ-mi) ma*	22-33	24	1	‘deal face to face’
接收马 VVN	*(tɕi-ɕou) ma*	22-33	24	1	‘receive a horse’
赶快动 AAV	*(gø-k^ha) doŋ*	44-22	24	0	‘hastily move’
贩卖米 VVN	*(fa-ma) mɛi*	44-22	24	0	‘sell rice’
经常买 AAV	*(tɕaŋ-ʔi) ma*	42-31	24	0	‘buy every day’
喜欢马 VVN	*(sʐ-ɕø) ma*	42-31	24	0	‘like horses’
预先谈 AAV	*(vu-ɕi) da*	33-22	31	1	‘talk beforehand’
贩卖田 VVN	*(fa-ma) di*	44-22	31	1	‘sell the land’
大声谈 AAV	*(da-sɜŋ) da*	33-22	31	1	‘talk loudly’
贩卖床 VVN	*(fa-ma) juɔ*	44-22	31	1	‘sell a bed’
当面谈 AAV	*(tuɔ-mi) da*	22-33	31	0	‘talk face to face’
相信人 VVN	*(ɕi-saŋ) naŋ*	22-33	31	0	‘believe in a person’
干脆谈 AAV	*(kø-ts^hai) da*	22-33	31	0	‘simply talk’
接近人 VVN	*(tɕi-dʑaŋ) naŋ*	31-35	31	0	‘approach a person’
当面打 AAV	*(tuɔ-mi) tiɛ*	22-33	35	1	‘hit face to face’
接收狗 VVN	*(tɕi-ɕou) kau*	22-33	35	1	‘receive a dog’

干脆走 AAV	*(ky-ts^{h}ai) tsau*	22-33	35	1	‘simply walk’
接近狗 VVN	*(tchi-dzhang) kau*	31-35	35	1	‘approach a dog’
经常走 AAV	*(tchang-ʔi) tsau*	42-31	35	0	‘walk every day’
喜欢狗 VVN	*(si-chy) kau*	42-31	35	0	‘like dogs’
快点走 AAV	*(k^{h}a-ti) tsau*	44-22	35	0	‘walk quickly’
再次算 AAV	*(tsɛi-ts^{h}ʐ̩) sɵ*	42-31	42	0	‘count once more’
喜欢菜 VVN	*(sʐ̩-ɕø) ts^{h}ɛi*	42-31	42	1	‘like vegetables’
赶紧剁 AAV	*(kø-tɕaŋ) tou*	44-22	42	1	‘chop hurriedly’
贩卖报 VVN	*(fa-ma) pɜ*	44-22	42	1	‘sell a report’
当面教 AVV	*(tuɔ-mi) kuɔ*	31-35	42	1	‘teach face to face’
接近店 VVN	*(tɕi-dʑaŋ) ti*	31-35	42	0	‘approach a shop’
干脆喝 AAV	*(kø-ts^{h}ai) hɜ*	22-33	42	0	‘simply drink’
接收报 VVN	*(tɕi-ɕou) pɜ*	22-33	42	0	‘receive a report’

Appendix 7.1

Stimulus materials for experiment 1 in Chapter 7. Questions prompting subject focus (SF), VP focus (VPF), or object focus (OF) are transcribed below as they were given by the speaker with whom they were recorded (see section 7.2.1). Picture references to the QUIS Reference Manual (Skopeteas et al. 2006).

(1) **woman – hit – man** (QUIS p.101, sheet 1)

SF	*(ʔa) ni*	*naŋ*	*ʔa*	*tiɛ*	*ki*	*kai*	*nɵ?*
	which	person	Q	hit	this	CL	male
	‘Who hits this man?’						
VPF	*kai*	*nø*	*zẓ ta*	*tsou*	*ʔa tɕi zɜ kɵ?*		
	CL	female	ASP	do	which thing		
	‘What is the woman doing?’						
OF	*kai*	*nø*	*zẓ ta*	*tiɛ*	*ʔa ni*	*naŋ*	*(ʔa)?*
	CL	female	ASP	hit	which	person	Q
	‘Who does the woman hit?’						

(2) **girl – hit – girl (- with hammer)** (QUIS p.101, sheet 1)

SF	*ʔa ni*	*naŋ*	*(ʔa)*	*jou*	*ki kai luɔ tou*		*zẓ ta*	*k^huo*
	which	person	Q	there.is	this CL	hammer	ASP	smack
	kai	*nø*	*mai mai?*					
	CL	female	little.child					
	‘Who is hitting the girl with a hammer?’							
VPF	*kai nø*	*mai mai*		*zẓ ta*	*tɕø*	*(a) ni?*		
	CL female	little.child		ASP	do	what		
	‘What is the girl doing?’							
OF	*kai*	*nø*	*zẓ ta*	*tiɛ*	*ʔa ni*	*naŋ*	*(ʔa)?*	
	CL	female	ASP	hit	which	person	Q	
	‘Who is the girl hitting?’							

(3) **girl – hit – car (- with hammer)** (QUIS p.102, sheet 4)

SF	*ʔa ni*	*naŋ*	*(ʔa)*	*jou k^huɔ luɔ tou*		*tchuɔ lɜ k^hu*		*ts^hu?*
	which	person	Q	there.is CL	hammer	hit	CL	car
	‘Who hits the car with a hammer?’							

VPF *kai nø mai mai zẓ ta tɕø (a) ni?*
CL female little.child ASP do what
'What is the girl doing?'
OF *kai nø mai mai zẓ ta t^{h}ɛi ʔa ni mɵ zẓ?*
CL female little.child ASP hit which thing
'What is the girl hitting?'

(4) **woman – hit – cow** (QUIS p. 79, item 2 picture 2)
SF *(ʔa) ni naŋ (ʔa) zẓ ta tiɛ ki noŋ ŋau?*
which person Q ASP hit this CL cow
'Who is hitting the cow?'
VPF *kai nø zẓ ta tɕø (a) ni?*
CL female ASP do what
'What is the woman doing?'
OF *kai nø zẓ ta tiɛ ʔa ni mɵ zẓ?*
CL female ASP hit which thing
'What is the woman hitting?'

(5) **bike – hit – woman** (QUIS p. 66, item 4 picture 2)
SF *ʔa ni moe zi bɛi kai nø tɕuɔ tɕaŋ?*
which thing to CL female hit
'What hits the woman?'
VPF *kai nø hu siɛ ʔa ni zɜ kɵ?*
CL female is affected by which thing
'What happens to the woman?'
OF *ki pu t^{h}ai da ts^{h}u ta tɕø (a) ni?*
this CL bike ASP do what
'Who does the bike hit?'

(6) **man – kick – chair** (QUIS p.101, sheet 1)
SF *ʔa ni naŋ ʔa t^{h}ɛi $tɕ^{h}$a taŋ?*
which person Q kicks CL chair
'Who kicks the chair?'
VPF *kai nɵ zẓ ta tɕø (a) ni?*
CL male ASP do what
'What is the man doing?'

OF *kai nɵ zẓ ta t^{h}ɛi ʔa ni mɵ zẓ?*
CL male ASP kick which thing
'What is the man kicking?'

(7) **girl – kick – boy** (QUIS p. 202, item 2)
SF *(ʔa) ni naŋ (ʔa) (zẓ t)a t^{h}ɛi ki kai nɵ mai mai?*
which person Q ASP kick this CL male little.child
'Who is kicking the boy?'
VPF *kai nø mai mai zẓ ta tɕø (a) ni?*
CL female little.child ASP do what
' What is the girl doing?'
OF *kai nø mai mai zẓ ta t^{h}ɛi ʔa ni naŋ (ʔa)?*
CL female little.child ASP kick which person Q
'Who is the girl kicking?'

(8) **boy – kick – man** (QUIS p. 41, item 1 picture 2)
SF *ʔa ni naŋ ʔa t^{h}ɛi ki kai nɵ ki?*
which person Q kicks this CL male
'Who kicks the man?'
VPF *kai nɵ mai maizẓ ta tɕø (a) ni?*
CL male little.child ASP do what
'What is the boy doing?'
OF *kai nɵ mai maizẓ ta t^{h}ɛi ʔa ni naŋ (ʔa)?*
CL male little.child ASP kick which person Q
'Who is the boy kicking?'

(9) **woman – kick – ball** (QUIS p. 36 item 3 picture 3)
SF *(ʔa) ni naŋ ʔa t^{h}ɛi ki kai tɕau?*
which person Q kick this CL ball
'Who kicks the ball?'
VPF *kai nø zẓ ta tɕø (a) ni?*
CL female ASP do what
'What is the woman doing?'
OF *kai nø zẓ ta t^{h}ɛi ʔa ni mɵ zẓ?*
CL female ASP kick which thing
'What is the woman kicking?'

(10) **man – push – car** (QUIS p.101, sheet 1)

SF	*(ʔa) ni*	*naŋ*	*(ʔa)*	*zʐ ta*	*tɜ*	*ki pu ts^{h}u?*
	which	person	Q	ASP	push	this CL car
	‘Who is pushing the car?’					
VPF	*kai nɵ*		*zʐ ta*	*tɕø*	*(a) ni?*	
	CL male		ASP	do	what	
	‘What is the man doing?’					
OF	*kai nɵ*		*zʐ ta*	*tɜ*	*ʔa ni*	*mɵ zʐ?*
	CL male		ASP	push	which	thing
	‘What is the man pushing?’					

(11) **man – push – (other) man** (QUIS p.101, sheet 2)

SF	*(ʔa) ni*	*naŋ*	*(ʔa)*	*zʐ ta*	*tɜ*	*ki kai nɵ?*
	which	person	Q	ASP	push	this CL man
	‘Who is pushing this man?’					
VPF	*kai nɵ*		*zʐ ta*	*tɕø*	*(a) ni?*	
	CL male		ASP	do	what	
	‘What is the man doing?’					
OF	*kai nɵ*		*ha*	*ʔa ni*	*naŋ*	*ʔa*	*tɜ?*
	CL male		OBJ	which	person	Q	push
	‘Who does the man push?’						

(12) **woman – push – girl** (QUIS p. 202, item 4)

SF	*ʔa (n)i*	*naŋ*	*ʔa*	*tɜ*	*kai nø*	*mai mai?*
	which	person	Q	pushes	CL female	little.child
	‘Who pushes the girl?’					
VPF	*kai nø*		*zʐ ta*	*tɕø*	*(a) ni?*	
	CL female		ASP	do	what	
	‘What is the woman doing?’					
OF	*kai nø*		*zʐ ta*	*tɜ*	*ʔa ni*	*naŋ*	*(ʔa)?*
	CL female		ASP	push	which	person	Q
	‘Who is the woman pushing?’						

(13) **boy – pull – man** (QUIS p. 202, item 3)

SF	*ʔa ni*	*naŋ*	*(ʔa)*	*zʐ ta*	*la*	*kai nɵ?*
	which	person	Q	ASP	pull	CL male
	‘Who is pulling the man?’					

VPF	*kai nɵ mai maizʐ ta*	*tɕø*	*(a) ni?*			
	CL male little.child	ASP	do	what		
	'What is the boy doing?'					
OF	*kai nɵ mai maizʐ ta*	*la*	*ʔa ni*	*naŋ*	*(ʔa)?*	
	CL male little.child	ASP	pull	which	person	Q
	'Who is the boy pulling?'					

(14) **girl – pull – chair** (QUIS p. 144, item 5s)

SF	*(ʔa) ni naŋ*	*(ʔa)*	*(zʐ) ta*	*la*	*ki tɕa taŋ?*	
	which person	Q	ASP	pull	this CL	chair
	'Who is pulling the chair?'					
VPF	*kai nø mai mai*		*zʐ ta*	*tɕø*	*(a) ni?*	
	CL female little.child		ASP	do	what	
	'What is the girl doing?'					
OF	*kai nø mai mai*		*zʐ ta*	*la*	*ʔa ni?*	
	CL female little.child		ASP	pull	what	
	'What is the girl pulling?'					

Practice Item:

(15) **boy – throw – ball** (QUIS p. 144, item 4s)

SF	*ʔa (n)i naŋ*	*ʔa*	*ta*	*ki kai tɕau?*		
	which person	Q	throws	this CL	ball	
	'Who throws the ball?'					
"Bad" answer:	*kai nɵ mai mai.*					
	CL male little.child					
	‚A boy.'					
"Good" answer:	*kai nɵ mai mai*	*zʐ ta*	*ta*	*ki kai tɕau.*		
	CL male little.child	ASP	throw	this CL	ball	
	'A boy is throwing the ball.'					
OF	*kai nɵ mai maizʐ ta*	*ta*	*ʔa ni*	*mɵ zʐ?*		
	CL male little.child	ASP	throw	which	thing	
	'What is the boy throwing?'					
"Bad" answer:	*ʔi kai tchau.*					
	NUM CL ball					
	‚A ball.'					
"Good" answer:	*kai nɵ mai mai*	*zʐ ta*	*ta*	*ʔi kai tɕau.*		
	CL male little.child	ASP	throw	NUM CL	ball	
	'The boy is throwing a ball.'					

Appendix 7.2

Welcome Message for experiment 1 in Chapter 7.

欢迎参加这个实验！
请仔细阅读实验步骤。 如果有问题，请问实验人。

这个实验需要您看图片回答问题。 请先看图片，然后仔细听问题。听到问题后再根据您看到的图片回答问题。

请用完整的句子，自然流畅的语气来回答问题。答完一个问题后，请点击空格键，进入下一张图片。您随时可以休息几分钟。

请点击空格键.

Welcome to participate in this experiment!
Please read the experimental instructions carefully. If you have any questions, please ask the experimenter.

This experiment lets you look at a picture and asks you to answer a question. Please look at the picture first, and then listen to the question carefully. After having heard the question, please answer it on basis of what you have seen in the picture.

Please use complete sentences when answering the questions, and speak in a natural and easy manner. After answering the question, please press the space bar for the next picture to appear. You may take a rest for a few minutes at any time.

Please press the space bar.

Appendix 7.3

Target sentences and precursor questions for experiment 2 in Chapter 7.
Focus conditions: BF = Broad focus, SF = Subject focus, VPF = VP focus, OF = Object focus

(1) **Target sentence:**

阿妈	正在	做	饭.
ʔa ba	*zɻ ta*	*tsou*	*va*
dad	ASP	make	rice

'Dad is preparing rice.'

Precursor questions:

BF	你	说	什么啊?			
	ni	*kuɔ*	*ʔa ni ʔa*			
	you	say	what Q			
	'What did you say?'					
SF	什么	人	啊	正在	做	饭?
	ʔa ni	*naŋ*	*ʔa*	*zɻ ta*	*tsou*	*va*
	which	person	Q	ASP	make	rice
	'Who is preparing rice?'					
VPF	阿妈	正在	作	什么啊?		
	ʔa ba	*zɻ ta*	*tsu*	*ʔa ni ʔa*		
	dad	ASP	do	what Q		
	'What is dad doing?'					
OF	阿妈	正在	做	什么啊?		
	ʔa ba	*zɻ ta*	*tsou*	*ʔa ni ʔa*		
	dad	ASP	make	what Q		
	'What is dad preparing?'					

(2) **Target sentence:**

阿爸	正在	买	东西.
ʔa ba	*zɻ ta*	*ma*	*mɵ zɻ*
dad	ASP	buy	groceries

'Dad is buying groceries.'

Precursor questions:

BF	你	说	什么啊？			
	ni	*kuɔ*	*ʔa ni ʔa*			
	you	say	what Q			
	‘What did you say?’					
SF	什么	人	啊	正在	买	东西？
	ʔa ni	*naŋ*	*ʔa*	*zɻ ta*	*ma*	*mɵ zɻ*
	which	person	Q	ASP	buy	groceries
	‘Who is buying groceries?’					
VPF	阿爸	正在	作	什么啊？		
	ʔa ba	*zɻ ta*	*tsu*	*ʔa ni ʔa*		
	dad	ASP	do	what Q		
	‘What is dad doing?’					
OF	阿爸	正在	买	什么啊？		
	ʔa ba	*zɻ ta*	*ma*	*ʔa ni ʔa*		
	dad	ASP	buy	what Q		
	‘What is dad buying?’					

(3) **Target sentence:**

阿妈	正在	抬	一个	男妹妹.
ʔa ma	*zɻ ta*	*tɛi*	*ʔi kai*	*nɵ mai mai*
mom	ASP	carry	NUM CL	male little.child

‘Mom is carrying a boy.’

Precursor questions:

BF	你	说	什么啊？				
	ni	*kuɔ*	*ʔa ni ʔa*				
	you	say	what Q				
	‘What did you say?’						
SF	什么	人	啊	正在	抬	一个	男妹妹？
	ʔa ni	*naŋ*	*ʔa*	*zɻ ta*	*tɛi*	*ʔi kai*	*nɵ mai mai*
	which	person	Q	ASP	carry	NUM CL	male little.child
	‘Who is carrying a boy?’						
VPF	阿妈	正在	作	什么啊？			
	ʔa ma	*zɻ ta*	*tsu*	*ʔa ni ʔa*			
	mom	ASP	do	what Q			
	‘What is mom doing?’						

OF 阿妈 正在 抬 什么人啊？
ʔa ma zẓ ta tɛi ʔa ni naŋ ʔa
mom ASP carry which person Q
'Who is mom carrying?'

(4) **Target sentence:**
阿妈 正在 拉 一个 女妹妹.
ʔa ma zẓ ta la ʔi kai nø mai mai
mom ASP drag NUM CL female little.child
'Mom is dragging a girl.'

Precursor questions:
BF 你 说 什么啊？
ni kuɔ ʔa ni ʔa
you say what Q
'What did you say?'
SF 什么 人 啊 正在 拉 一个 女妹妹？
ʔa ni naŋ ʔa zẓ ta la ʔi kai nø mai mai
which person Q ASP drag NUM CL female little.child
'Who is dragging a girl?'
VPF 阿妈 正在 作 什么啊？
ʔa ma zẓ ta tsu ʔa ni ʔa
mom ASP do what Q
'What is mom doing?'
OF 阿妈 正在 拉 什么人啊？
ʔa ma zẓ ta la ʔa ni naŋ ʔa
mom ASP drag which person Q
'Who is mom dragging?'

(5) **Target sentence:**
一个男 正在 吻 一个女.
ʔi kai nɵ zẓ ta vaŋ ʔi kai nø
NUM CL male ASP kiss NUM CL female
'A man is kissing a woman.'

Precursor questions:

BF	你	说	什么啊？		
	ni	*kuɔ*	*ʔa ni ʔa*		
	you	say	what Q		
	‘What did you say?’				
SF	什么	人	啊	正在	吻	一个女？
	ʔa ni	*naŋ*	*ʔa*	*zẓ ta*	*vaŋ*	*ʔi kai nø*
	which	person	Q	ASP	kiss	NUM CL female
	‘Who is kissing a woman?’					
VPF	一个男		正在	作	什么啊？	
	ʔi kai nɵ		*zẓ ta*	*tsu*	*ʔa ni ʔa*	
	NUM CL male		ASP	do	what Q	
	‘What is the man doing?’					
OF	一个男		正在	吻	什么人啊？	
	ʔi kai nɵ		*zẓ ta*	*vaŋ*	*ʔa ni naŋ ʔa*	
	NUM CL male		ASP	kiss	which person Q	
	‘Who is the man kissing?’					

(6) **Target sentence:**

一个女	正在	打	一个牛.
ʔi kai nø	*zẓ ta*	*tiɛ*	*ʔi doŋ ŋau*
NUM CL female	ASP	hit	NUM CL cow

‘A woman is hitting a cow.’

Precursor questions:

BF	你	说	什么啊？		
	ni	*kuɔ*	*ʔa ni ʔa*		
	you	say	what Q		
	‘What did you say?’				
SF	什么	人	啊	正在	打	一个牛？
	ʔa ni	*naŋ*	*ʔa*	*zẓ ta*	*tiɛ*	*ʔi doŋ ŋau*
	which	person	Q	ASP	hit	NUM CL cow
	‘Who is hitting a cow?’					
VPF	一个女		正在	作	什么啊？	
	ʔi kai nø		*zẓ ta*	*tsu*	*ʔa ni ʔa*	
	NUM CL female		ASP	do	what Q	
	‘What is the woman doing?’					

OF 一个女 正在 打 什么啊？
ʔi kai nø *zɻ ta* *tiɛ* *ʔa ni ʔa*
NUM CL female ASP hit what Q
'What is the woman hitting?'

(7) **Target sentence:**
一个男 正在 筑 一张桌子.
ʔi kai nɵ *zɻ ta* *tɕou* *ʔi tɕi dʐu*
NUM CL male ASP build NUM CL table
'A man is building a table.'

Precursor questions:
BF 你 说 什么啊？
ni *kuɔ* *ʔa ni ʔa*
you say what Q
'What did you say?'
SF 什么 人 啊 正在 筑 一张桌子？
ʔa ni *naŋ* *ʔa* *zɻ ta* *tɕou* *ʔi tɕi dʐu*
which person Q ASP build NUM CL table
'Who is building a table?'
VPF 一个男 正在 作 什么啊？
ʔi kai nɵ *zɻ ta* *tsu* *ʔa ni ʔa*
NUM CL male ASP do what Q
'What is the man doing?'
OF 一个男 正在 筑 什么啊？
ʔi kai nɵ *zɻ ta* *tɕou* *ʔa ni ʔa*
NUM CL male ASP build what Q
'What is the man building?'

(8) **Target sentence:**
一个女 正在 喝 一杯茶.
ʔi kai nø *zɻ ta* *ha* *ʔi bai dzu*
NUM CL female ASP drink NUM CL tea
'A woman is drinking a cup of tea.'

Precursor questions:

BF	你	说	什么啊？			
	ni	*kuɔ*	*ʔa ni ʔa*			
	you	say	what Q			
	‘What did you say?’					
SF	什么	人	啊	正在	喝	一杯茶？
	ʔa ni	*naŋ*	*ʔa*	*zẓ ta*	*ha*	*ʔi bai dzu*
	which	person	Q	ASP	drink	NUM CL tea
	‘Who is drinking a cup of tea?’					
VPF	一个女		正在	作	什么啊？	
	ʔi kai nø		*zẓ ta*	*tsu*	*ʔa ni ʔa*	
	NUM CL female		ASP	do	what Q	
	‘What is the woman doing?’					
OF	一个女		正在	喝	什么啊？	
	ʔi kai nø		*zẓ ta*	*ha*	*ʔa ni ʔa*	
	NUM CL female		ASP	drink	what Q	
	‘What is the woman drinking?’					

(9) **Target sentence:**

一个	男妹妹	正在	拉	阿爸.
ʔi kai	*nɵ mai mai*	*zẓ ta*	*la*	*ʔa ba*
NUM CL	male little.child	ASP	drag	dad

‘A boy is dragging dad.’

Precursor questions:

BF	你	说	什么啊？			
	ni	*kuɔ*	*ʔa ni ʔa*			
	you	say	what Q			
	‘What did you say?’					
SF	什么	人	啊	正在	拉	阿爸？
	ʔa ni	*naŋ*	*ʔa*	*zẓ ta*	*la*	*ʔa ba*
	which	person	Q	ASP	drag	dad
	‘Who is dragging dad?’					
VPF	一个	男妹妹		正在	作	什么啊？
	ʔi kai	*nɵ mai mai*		*zẓ ta*	*tsu*	*ʔa ni ʔa*
	NUM CL	male little.child		ASP	do	what Q
	‘What is the boy doing?’					

OF 一个 男妹妹 正在 拉 什么人啊？
ʔi kai nɵ mai mai zẓ ta la ʔa ni naŋ ʔa
NUM CL male little.child ASP drag which person Q
'Who is the boy dragging?

(10) **Target sentence:**
一个 女妹妹 正在 骂 阿妈.
ʔi kai nø mai mai zẓ ta zu ʔa ma
NUM CL female little.child ASP scold mom
'A girl is scolding mom.'

Precursor questions:
BF 你 说 什么啊？
ni kuɔ ʔa ni ʔa
you say what Q
'What did you say?'
SF 什么 人 啊 正在 骂 阿妈？
ʔa ni naŋ ʔa zẓ ta zu ʔa ma
which person Q ASP scold mom
'Who is scolding mom?'
VPF 一个 女妹妹 正在 作 什么啊？
ʔi kai nø mai mai zẓ ta tsu ʔa ni ʔa
NUM CL female little.child ASP do what Q
'What is the girl doing?'
OF 一个 女妹妹 正在 骂 什么人啊？
ʔi kai nø mai mai zẓ ta zu ʔa ni naŋ ʔa
NUM CL female little.child ASP scold which person Q
'Who is the girl scolding?'

(11) **Target sentence:**
一个 男妹妹 正在 欺 一个 女妹妹.
ʔi kai nɵ mai mai zẓ ta ts^{h}ou ʔi kai nø mai mai
NUM CL male little.child ASP deceive NUM CL female little.child
'A boy is deceiving a girl.'

Precursor questions:

BF 你 说 什么啊？
ni *kuɔ* *ʔa ni ʔa*
you say what Q
'What did you say?'

SF 什么 人 啊 正在 欺 一个 女妹妹?
ʔa ni *naŋ* *ʔa* *zẓ ta* *ts^{h}ou* *ʔi kai* *nø mai mai*
which person Q ASP deceive NUM CL female little.child
'Who is deceiving a girl?'

VPF 一个 男妹妹 正在 作 什么啊？
ʔi kai *nɵ mai mai* *zẓ ta* *tsu* *ʔa ni ʔa*
NUM CL male little.child ASP do what Q
'What is the boy doing?'

OF 一个 男妹妹 正在 欺 什么人啊？
ʔi kai *nɵ mai mai* *zẓ ta* *ts^{h}ou* *ʔa ni naŋ ʔa*
NUM CL male little.child ASP deceive which person Q
'Who is the boy deceiving?'

(12) **Target sentence:**

一个 女妹妹 正在 会 一个 男妹妹.
ʔi kai *nø mai mai* *zẓ ta* *vai* *ʔi kai* *nɵ mai mai*
NUM CL female little.child ASP meet NUM CL male little.child
'A girl is meeting a boy.'

Precursor questions:

BF 你 说 什么啊？
ni *kuɔ* *ʔa ni ʔa*
you say what Q
'What did you say?'

SF 什么 人 啊 正在 会 一个 男妹妹？
ʔa ni *naŋ* *ʔa* *zẓ ta* *vai* *ʔi kai* *nɵ mai mai*
which person Q ASP meet NUM CL male little.child
'Who is meeting a boy?'

VPF 一个 女妹妹 正在 作 什么啊？
ʔi kai *nø mai mai* *zẓ ta* *tsu* *ʔa ni ʔa*
NUM CL female little.child ASP do what Q
'What is the girl doing?'

OF	一个　女妹妹	正在 会　什么人啊？
	ʔɪ̀ kai　nø mai mai	*zẓ ta vai　ʔa ni naŋ ʔa*
	NUM CL female little.child	ASP meet which person Q
	‘Who is the girl meeting?’	

Curriculum vitae

Franziska Scholz was born in Berlin in the former German Democratic Republic on 10 january 1983. She attended primary and secondary school in Berlin, and graduated from the Heinrich-von-Kleist Gymnasium in 2002. In the same year, she took up her studies of Germanic Linguistics, Political science, and Modern history at the Humboldt-University in Berlin. During her studies, she spent one semester in 2005/2006 at the University of Manchester, where she took classes in English Language and Linguistics as well as in General Linguistics. In 2008, she obtained her Magister degree (cum laude) at the Humboldt-University in Berlin, and proceeded to join the research project “An experimental approach to the interaction of tone sandhi and focus expression in six dialects of Chinese” of dr. Yiya Chen at the Leiden University Centre for Linguistics in the Netherlands as an *assistent in opleiding* (doctoral researcher). This dissertation is the result of her research in this project.