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

Scholz, F.

# 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.