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

Scholz, F.

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]

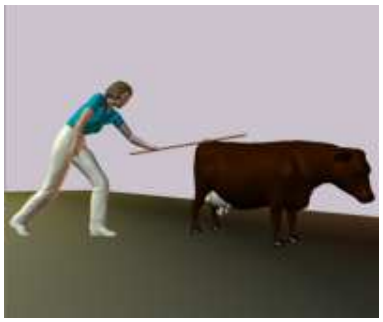

Figure 7.1: Example stimulus picture from the QUIS corpus (Skopeteas et al. 2006: 79, item 2 picture 2). Reprinted with permission of the SFB 632/ Malte Zimmermann.

(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

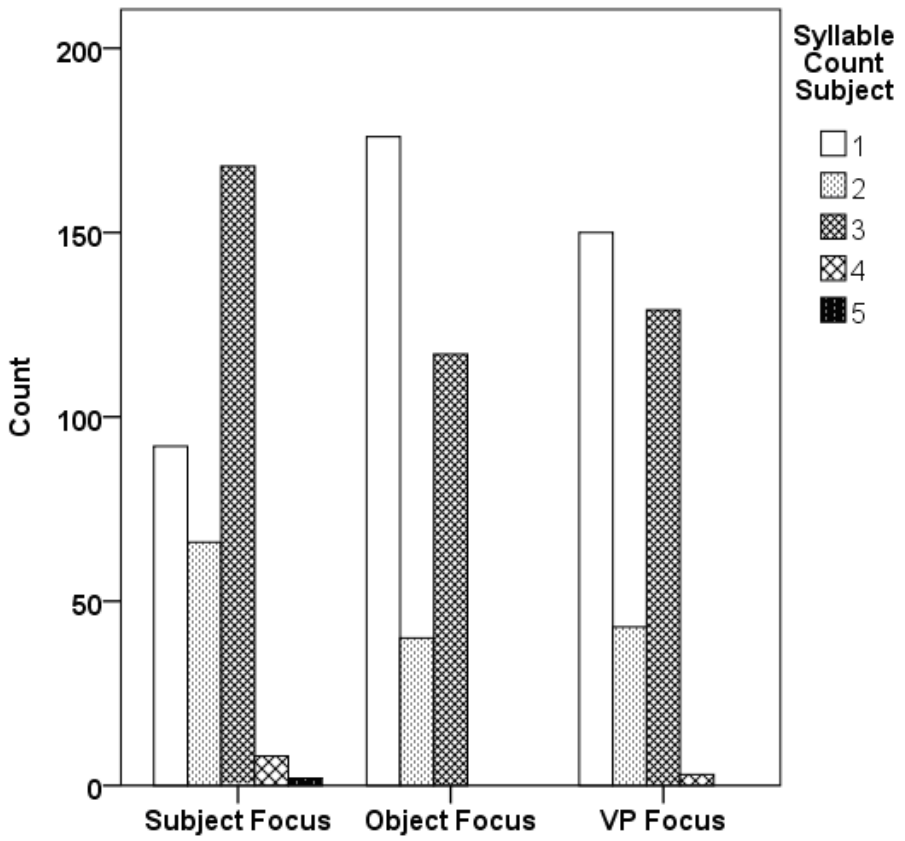

Figure 7.2: Count of instances of subject NPs with one to five syllables, broken down by focus condition.

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 [$\chi^2(10) = 41.75$, $p < 0.001$].

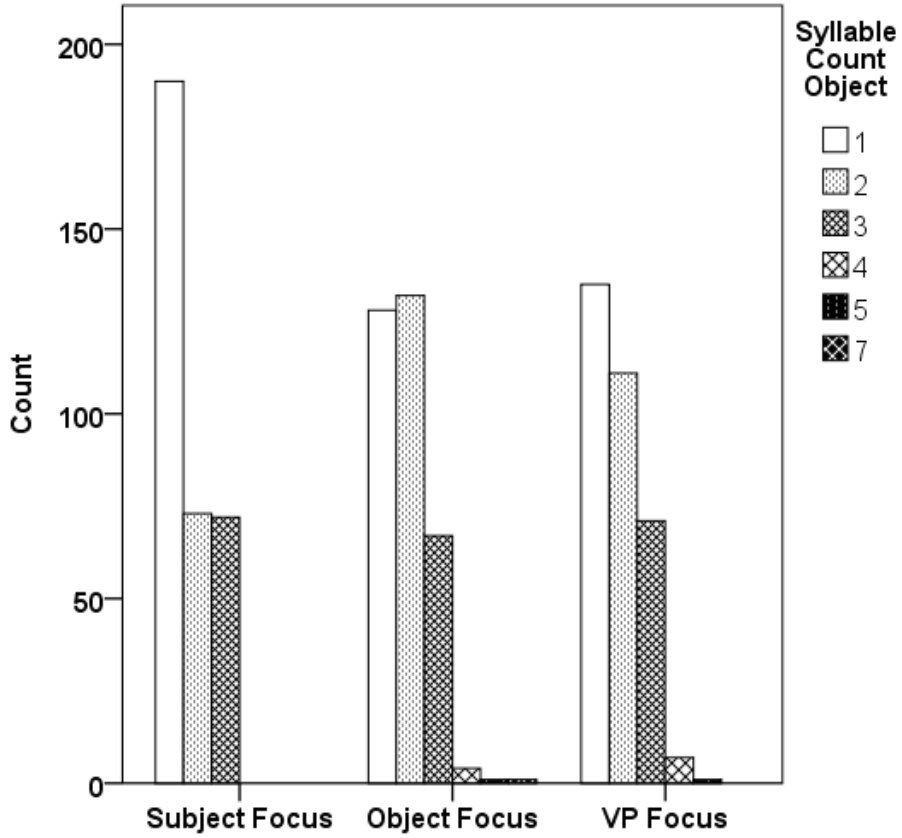

Figure 7.3: Count of instances of object NPs with 1-7 syllables, broken down by focus condition.

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

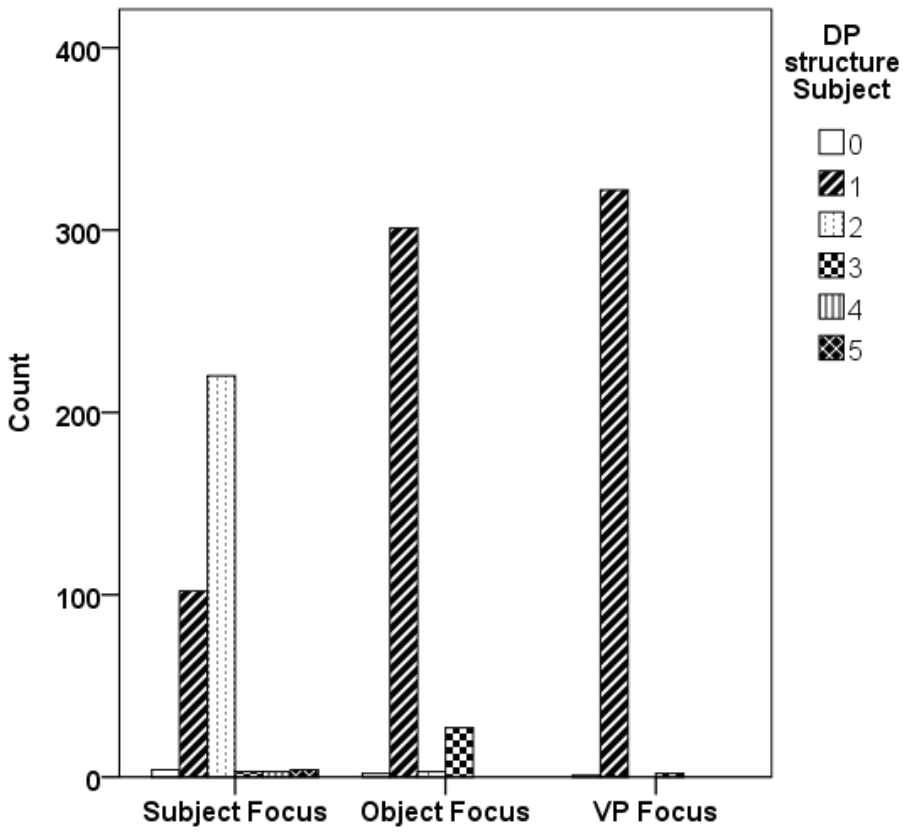

Figure 7.4: Count of instances of subject DPs with different degree of specificity, broken down by focus condition.

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 [$\chi^2(14) = 334.81$, $p < 0.001$].

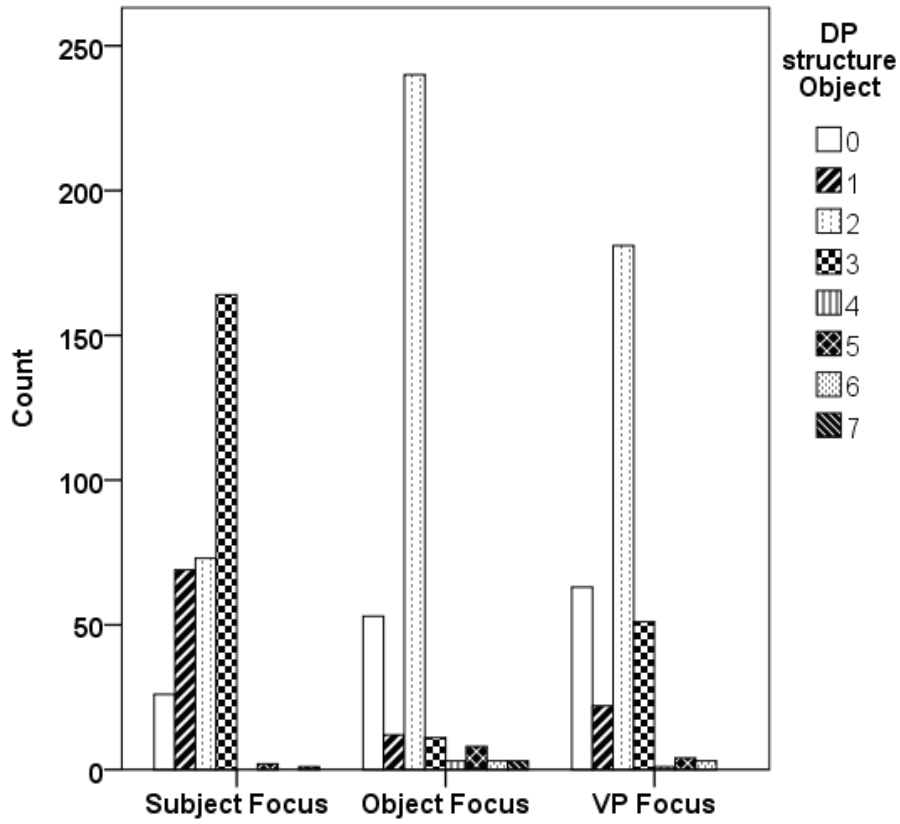

Figure 7.5: Count of instances of object DPs with different structures, broken down by focus condition.

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.

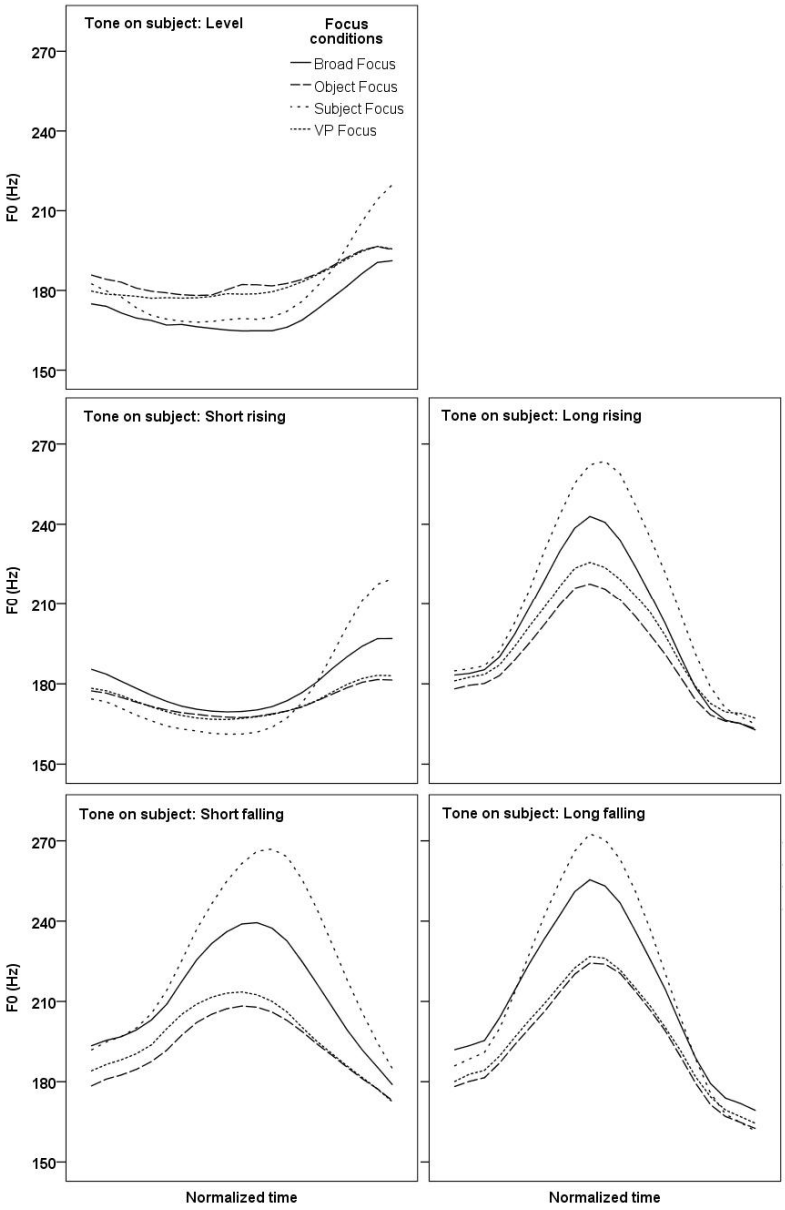

Figure 7.6: Averaged time-normalized tonal contours for subject referents.

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).

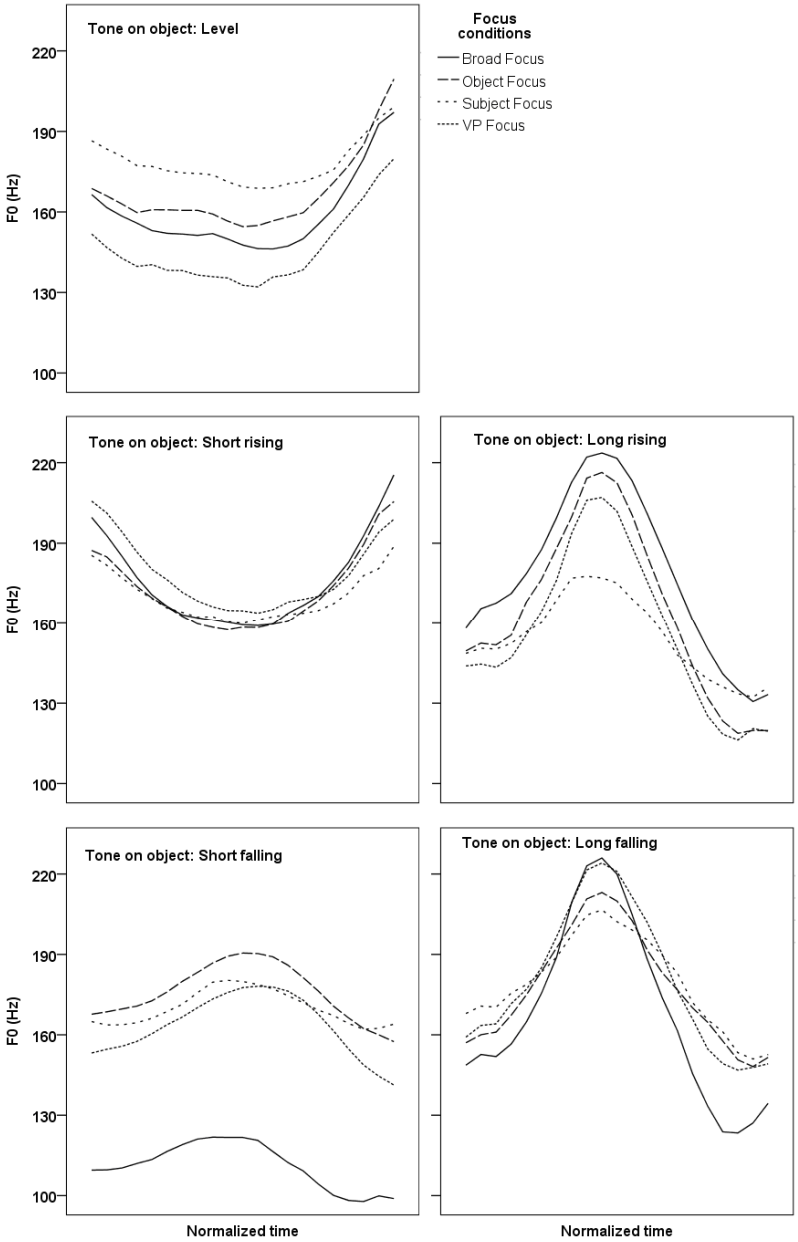

Figure 7.7: Averaged time-normalized tonal contours for object referents.

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.

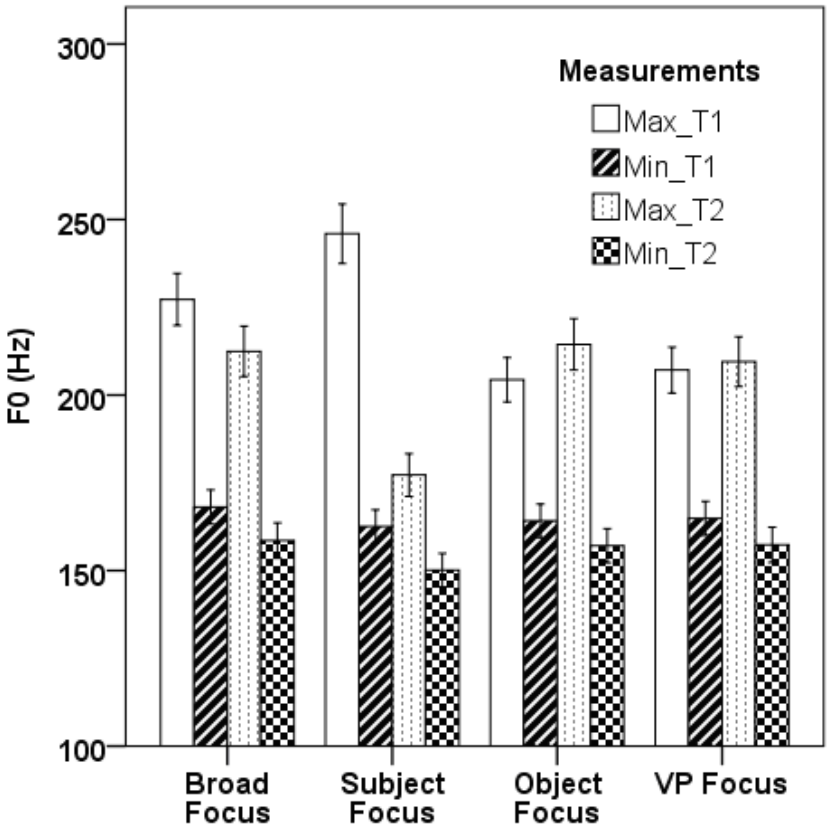

Figure 7.8: Averaged F_0 maximum and F_0 minimum values for the subject (T1) and object (T2) constituent, broken down by focus condition.

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.

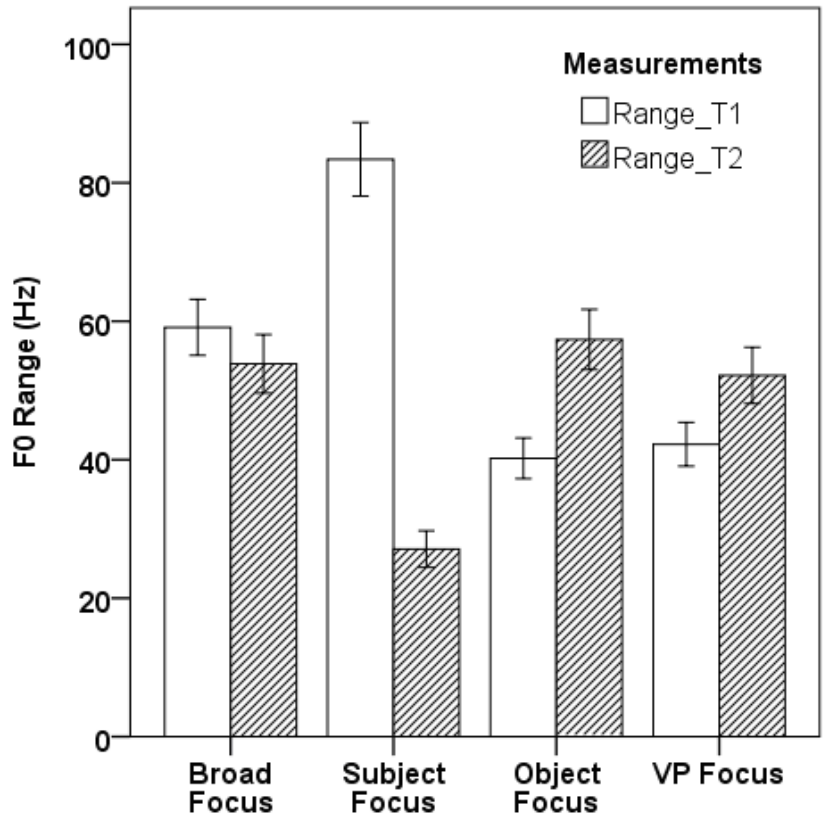

Figure 7.9: Averaged F_0 range values for the subject (T1) and object (T2) constituent, broken down by focus condition.

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.

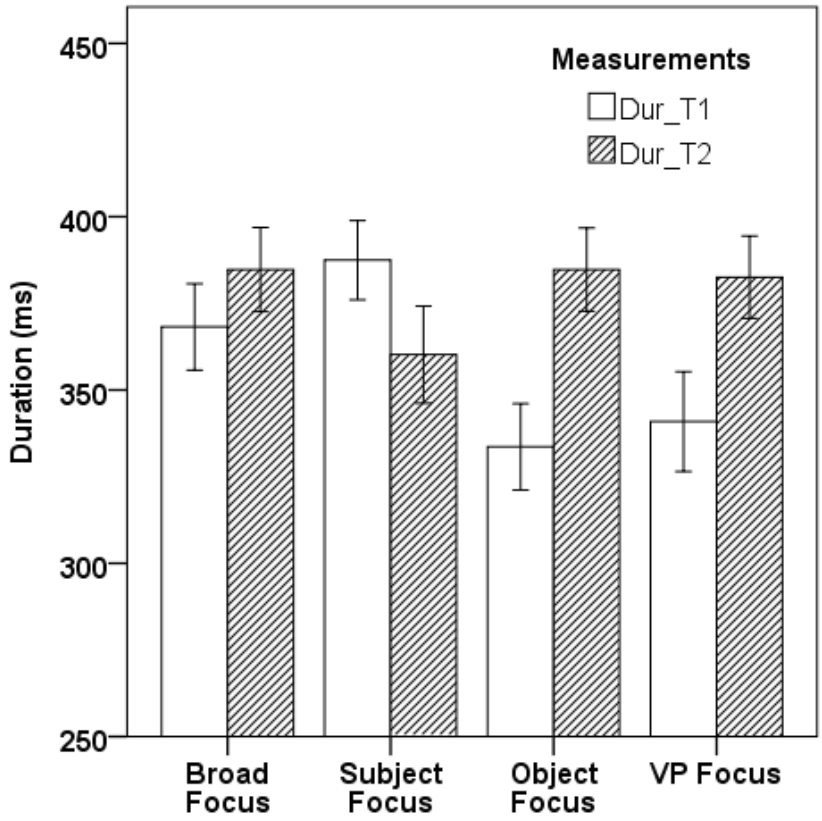

Figure 7.10: Averaged duration values for the subject (T1) and object (T2) constituents, broken down by focus condition.

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.