UCLA
UCLA Electronic Theses and Dissertations

Title
An Autosegmental-Metrical Model of Shanghainese Tone and Intonation

Permalink
https://escholarship.org/uc/item/5hm0n8b7

Author
Roberts, Brice David

Publication Date
2020

Supplemental Material
https://escholarship.org/uc/item/5hm0n8b7#supplemental

Peer reviewed|Thesis/dissertation

eScholarship.org Powered by the California Digital Library
University of California

UNIVERSITY OF CALIFORNIA
Los Angeles

An Autosegmental Metrical Model of Shanghainese Tone and Intonation

A dissertation submitted in partial satisfaction of the
requirements for the degree Doctor of Philosophy
in Linguistics

by

Brice David Roberts

2020

© Copyright by

Brice David Roberts

2020

ABSTRACT OF THE DISSERTATION

An Autosegmental-Metrical Model
of Shanghainese Tone and Intonation

by

Brice David Roberts
Doctor of Philosophy in Linguistics
University of California, Los Angeles, 2020
Professor Sun-Ah Jun, Chair

This dissertation presents a model of Shanghainese lexical tone and intonation based in the Autosegmental-Metrical framework and develops an annotation system for prosodic events in the language, known as Shanghainese Tones and Break Indices Labeling, or Sh_ToBI. Full-sentence phonetic data from 21 Shanghainese speakers (born 1937-1975) were analyzed.

Instead of a syllable tone language with left-dominant sandhi, Shanghainese is analyzed here as a lexical pitch accent language, with three levels of phrasing above the syllable. The lowest level of phrasing is the accentual phrase, which is the domain of the three contrastive pitch accents, H*, L*+H, and L*. These pitch accents are paired with one of two AP-final boundary tones: La/L:a or LHa. La/L:a varies freely between a single low target (La) and a low plateau (L:a), and co-occurs with H* and L*+H. LHa is a sharp rising boundary tone which accompanies L*. AP boundary tones always accompany simple pitch accents (H* and L*), while they only appear after bi-tonal L*+H in APs longer than two syllables. AP-initial tone targets, H* and L*, are prominent and are always the local pitch maximum or minimum, respectively.

Above the AP is the intermediate phrase (ip), identifiable only by its phrase-final segmental lengthening, following pitch reset, and a lack of following silence. The highest level is the intonational phrase (IP) marked with initial pitch range expansion, and one of three IP-final boundary tones (H%, L%, or toneless %). H% is used in particle-final yes/no questions, interacting variably with preceding tones. L% is used in corrective or other emotionally marked statements, creating contours with preceding tones. % is toneless, and is used in most questions and all unmarked statements. It is recognizable via segmental lengthening, pitch range compression, and a following silence.

Beyond this re-analysis of Shanghainese tone and phrasing, the dissertation finds intonational differences between two kinds of focus (general narrow and corrective), and explores the application of tonal reduction (also called ‘right-dominant sandhi’) to both monosyllabic and disyllabic APs. Finally, a break index system for labeling the perceived amount of juncture between syllables was developed.

The dissertation of Brice David Roberts is approved.

Bruce Hayes

Jie Zhang

Kie Zuraw

Sun-Ah Jun, Committee Chair

University of California, Los Angeles

2020

Dedicated to:

My father—David K. Roberts

My mother—Janet A. Roberts

Table of Contents

List of Figures

List of Tables

Acknowledgements

I admit that I have been thinking about this section of my dissertation more than is reasonable, as I knew I would want to give people proper thanks for their hand in my life and work. Of course, now that the time has come to put it into words, I find it difficult to write. This is simply due to the number of people that have aided me and the amount of support I have received, intentional or otherwise. I know that the following will still leave me unsatisfied with my expression of thanks, but it is the best I could do.

First is my advisor Sun-Ah Jun, who has overseen my academic career for the last six years. She has provided me not only with her knowledge and expertise concerning prosody, but also with moral support and encouragement along what was sometimes a difficult path to completing my dissertation. I must also thank my other committee members for their help, both academic and otherwise—Bruce Hayes for his ability to point out the claims I sometimes did not realize I was making (correctly or incorrectly), and for his monthly singings, now missed during the pandemic; Jie Zhang for sharing his encyclopedic knowledge of dialectal Chinese phonology and phonetics, and for his kind encouragement of my analysis; and Kie Zuraw for her classes and comments teaching me about the reality of phonological knowledge and how to be a better writer (though it remains to be seen how much of the latter has stuck with me). Above all, I must thank my committee for their patience and understanding while leading me through this process. In general, I feel very grateful to the entire UCLA Linguistics Department, for creating an environment in which I felt safe and supported in my career there, regardless of any struggles I have had along the way. Having other friends in other graduate programs around the country, I

appreciate just how rare a thing our department is, and I will always cherish my time spent with colleagues at spectrogram reading, Chinese lunches, or Georgian choir.

My earlier academic career was also marked by amazing educators. At Reed, I had the privilege of being taught by Alexei Ditter, Pyotr Gibas, Jing Jiang, and Hyong Rhew in the Chinese department and in Linguistics, Matt Pearson, Kara Becker, and my advisor Sameer ud Dowla Khan. I remember my time at Reed very fondly, in no small part due to the relationships with professors I had there.

I must also thank all those who were involved in the actual production of this specific work. I am greatly indebted to my speakers, who willingly met with a strange, tall American speaking subpar Mandarin in order to say countless bizarre sentences ad nauseam. Without their patience with me and passion for Shanghainese, this dissertation would truly not exist. My research assistant, Jessica Xinyi Wang, was also instrumental in completing this work—her native speaker judgments and easy-going nature were truly a godsend while in the throes of analyzing the language, and having her present in Shanghai for recordings made sure that the entire process went more smoothly than I could have hoped. The Sh_ToBI labeling system below would not look the same without her input. I must also thank the graduate students at the Fudan University Department of Chinese Language and Literature for welcoming me into their space and helping me find speakers. The help of Gary Yifei Zheng in particular was greatly appreciated—the participation of many speakers in this study can be directly attributed to his efforts to help with a project he had not heard of until a month prior. I am still touched by the amount of work he took on to help me.

Friends also had no small part in getting me to this point. I would be remiss to not mention my friends in the department itself—their friendship and curiosity has proven to be a

inspiration for me over the years. I also am always amazed by the strength of my friendships from college and even high school; without these people, I would not have had the escape from academic life that I desperately needed from time to time: John Troup, Julian Haft, Willamae Boling, Noah Atchison, Joe Deng, Rebecca 'RR' Richman, Sarah Evans, Andrew Choi, Drake Myers, and Evan Myers.

One friend in particular, Edita Cao, and her mother, Mei Jiang, deserve special thanks. These two women were my first introduction to Shanghainese, and sparked an interest that will last the rest of my life. It was only by becoming friends with Edie in high school German that I had the chance to encounter one of the most fascinating languages (in my opinion) and make one of the most enduring friendships of my life. As my only long-time friend enduring grad school at the same time as me, her support has been invaluable, and I find myself again and again thankful for her continued friendship.

Lastly, I could not have done any of this without the support of my family. No words can adequately express the strength that I have gotten from my mother, father, and brother over the years. Their consistent support and unlimited kindness has meant everything to me, even if from farther away than any of us would like. To my husband, Dwayne, I also extend my greatest thanks and love. These last six years have not always been easy for me, and thus were not particularly easy for you either, but we have found joy and happiness together, and I am forever grateful for that. I deeply appreciate everything you do to care for me, check in with me, make me laugh, and in general enjoy life.

Again, thank you, thank you, thank you—to everyone.

Vita

Education

University of California, Los Angeles

M.A.	Linguistics	June 2016
	Thesis	*A Preliminary Model of Shanghai Wu Intonational Phonology*
	Chair:	Sun-Ah Jun

Reed College

B.A.	Linguistics	2014
	Allied field:	Chinese
	Thesis:	*Searching for Stress in Shanghainese Tone Domains*
	Advisor:	Sameer ud Dowla Khan

Publications and Presentations (* indicates the presenter for a talk.)

Roberts, B. (July 2019) *A Model of Shanghai Wu Intonation*. An invited talk at the Department of Chinese Language and Literature, Fudan University. Shanghai, China.

Mantenuto, I.* & B. Roberts (January 2018) *Pronouns in San Sebastián del Monte Mixtec*. A talk given at the Society for the Study of the Indigenous Languages of the Americas. Salt Lake City, UT.

Mantenuto, I. & B. Roberts* (January 2018) *The Morphophonology of Aspect in San Sebastián del Monte Mixtec*. A talk given at the Society for the Study of the Indigenous Languages of the Americas. Salt Lake City, UT.

Roberts, B. (January 2017) *A Model of Shanghai Wu Intonational Phonology*. A poster at the Linguistic Society of America, Austin, TX.

Roberts, B. (2016) *A Preliminary Model of Shanghai Wu Intonational Phonology*. MA Thesis, UCLA.

1. Introduction to Dissertation

Shanghainese (also called Urban Shanghainese, Shanghai Wu, Shanghai dialect, and Shanghai, among others), which is a Chinese language in the Wu subfamily, has been the subject of much phonological and phonetic literature over the last century due mainly to its peculiar tone system (Chao 1928, Sherard 1972, Xu et al. 1986, Cao & Maddieson 1992, Gao 2015, and others). While many languages make lexical distinctions with tone, Shanghainese and its Wu cousins are known for their tendency to spread the lexical tone of some domain-initial syllable over the entire domain, deleting non-domain-initial tonal information to varying degrees, in a process known as tone sandhi. This makes it unlike the more well-known Chinese languages, such as Mandarin and Cantonese, whose tones generally pronounced the same whether produced in isolation or as a part of a longer phrase. Most of the work done on this language focuses on the phonological motivation for this sandhi process, or on the phonetic implementation thereof, with less research focused on how these sandhi domains are determined and interact with each other.

Also of growing interest in the past 40 years is the study of intonational and prosodic systems around the world. Intonation is another use of pitch that, instead of carrying lexical meaning, carries post-lexical, pragmatic, or emotional meaning. While not all languages have lexical tone like the Chinese languages, all languages have intonation. With the development of the Autosegmental-Metrical (AM) framework (Pierrehumbert 1980, Beckman & Pierrehumbert 1986, Pierrehumbert & Beckman 1988, Ladd 1996/2008, emanating from work by Liberman 1975, Goldsmith 1976 and Bruce 1977), many more languages have had their intonational phonology investigated for the first time. The AM analysis generally assumes two contrastive post-lexical tonal targets, High and Low, whether they mark prominence or the edge of prosodic

constituents. Finally, the framework is concerned with relating these underlying tones to the surface realization of a pitch track via rules, constraints, or other generalizations.

While intonation is universal, it remains somewhat understudied in lexical tone languages, no doubt in part due to the difficulty in disentangling the effects of the lexical and post-lexical tones. While the AM analyses have been applied to the intonation systems of Mandarin (Peng et al. 2005) and Cantonese (Wong et al. 2005), no other Chinese dialects have been analyzed under this framework. Because of the peculiarities of Shanghainese, even in comparison to other Chinese dialects, it seems a natural choice for further exploration in the realm of intonational typology. Thus this dissertation has the following questions and goals in applying the AM model of intonational phonology to Shanghainese:

i. What is the best characterization of the lexical tonal distinctions in Shanghainese?
ii. What are the prosodic units forming the phrasal hierarchy, and how can they be recognized?
iii. What is the inventory of contrastive post-lexical tones?
iv. How do tones (lexical or post-lexical) interact?
v. Create an annotation scheme for Shanghainese tones and intonation.

2. *Background Literature*

2.1.Linguistic Uses of Pitch

Pitch is the name for the percept of the lowest frequency in a complex sound wave, also known as the fundamental frequency, or f0. Pitch is used in all languages to convey information, both linguistic and non-linguistic. Of the linguistic uses of pitch, there is a main division between two types: lexical pitch and non-lexical pitch. Lexical uses of pitch, known as lexical tone, is a phenomenon whereby a language makes lexical distinctions by changing tones; e.g. Mandarin *mā* produced with a high-level tone means 'mother,' while *mǎ* with a low-dipping tone means 'horse.'

Intonation on the other hand, includes post-lexical tone distinctions as well as the rhythmic effects associated with phrasing and relative strength. Post-lexical tone refers to the manipulation of f0 used to convey both sentence-level linguistic (e.g. question vs. statement, syntactic constituency, focus) and non-linguistic (e.g. emotional state, fatigue) information over the course of an utterance. It is important to note that while not all languages have lexical tone (e.g. English, Georgian, Korean), all languages, even if tonal, have intonation. For an English example, compare the pitch over the following sentences:

1) a. The museum is closed.
   b. The museum is CLOSED?!

The first, with a generally falling pitch from the stressed syllable of 'mus*e*um' to the end of the sentence, is a simple statement. In the second, a question of disbelief, the pitch falls from the beginning of the sentence to 'CLOSED' which features an abrupt rise. While the amount of

incredulity may be emotional and thus non-linguistic, the change from statement to question is a categorical linguistic effect of the intonational contour. Notice that the meaning of ‘closed’ does not change dependent on its pitch, so even though there is a linguistic change signaled by the tones used, it is clearly post-lexical, not lexical tone.

2.2.Study of Intonation and the Autosegmental-Metrical Framework

Currently, one of the most widely-used frameworks in investigating intonation across languages is the Autosegmental-Metrical (AM) model. This model has its origins in multiple studies in the past 60 years, and has been used to describe and analyze many different intonational systems. In order to give a sense of the development of the theory, this section will outline the many influences on the framework, highlighting the key assumptions made by various researchers.

The first part of the framework’s name, ‘autosegmental,’ has its origins outside of the study of intonation directly. Autosegmentalism was developed by Goldsmith (1976) in order to analyze the phonological behavior of *lexical* tones in African languages, where abstract contrastive tonal targets can interact independently of the segmental string to which they were associated. The ‘metrical’ aspect of the theory comes from metrics, or the study of the relative prominence between phonological units (of any size) in an utterance and the way linguistic elements align to these prominences (Liberman & Prince 1977, Hayes 1983). Prominence in a phonological sense indicates some sort of ‘strength’ that can be realized phonetically in various ways, as will be described below. The phonological units that metrics is concerned with can be

both below the word (e.g. syllable and mora) and above the word (e.g. phonological word, phonological phrase).

Turning specifically to the study of intonation, current approaches are based in phonological analysis of phonetic data, which finds its origin in the Institute for Perception Research (IPO), which conducted intonational research generally between 1965 and 1995. They originally began with work on Dutch and eventually developed a general theory of intonational structure (‘t Hart et al. 1990). IPO approached intonation with two critical assumptions: 1) that sentential intonation is made up of a linear sequence of discrete and contrastive elements, and 2) that these discrete elements have a defined phonetic behavior (‘t Hart & Collier 1975). The first assumption echoes autosegmentalism, but their second assumption seems more specific to intonation. Rather than seeking phonetic correlates of specific pragmatic meanings over the whole sentence, the group identified consistent intonational events in order to categorize them referencing only their phonetic forms, not their function. IPO researchers also assumed that intonational tones are generally reducible to relatively high (H) targets and relatively low (L) tonal targets, and that some of these targets were ‘prominence-lending’ while others were not. The current autosegmental-metrical approach generally holds on to these assumptions.

Many researchers have specifically contributed to the development of the AM framework, like Liberman (1975), Bruce (1977), and Pierrehumbert (1980). In the decades since these dissertations, there has been a group of assumptions that has come to form the basis of how AM researchers approach analyzing a given language’s intonational phonology. First among these is that post-lexical tones behave independently of the words they are associated to—namely, ‘text’ and ‘tune’ are separate entities (Liberman 1975). The post-lexical tones are assumed to be arranged linearly, with only High and Low targets. Additionally, a major

distinction was made in the kinds of post-lexical tones that exist: pitch accents versus boundary tones. Pitch accents are the tonal events that generally mark some head, and thus generally align to prominent syllables, while boundary tones mark the edge of some phonological domain. While some languages have been found to lack pitch accents entirely (e.g. Korean; Jun 1993, 1995), all languages have boundary tones. Over syllables with no assigned tone, AM models assume pitch interpolation from the previous tone target to the next, ignoring any minor effects that segments may have on local pitch realization (also called 'micro-prosody').

In identifying the boundary tones of a given language, it is imperative that an AM analysis determine what boundary is actually being marked. Thus, there must be investigation into the number and kind of phonological phrase types that the language has, generally organized into a prosodic hierarchy. AM models of prosodic hierarchy hold to the assumption of the strict-layering hypothesis (Selkirk 1984, 1986; Nespor & Vogel 1986/2007), which constrains the kinds of phrasing allowed across languages. In its simplest form, it states that 1) every prosodic unit must dominate or contain at least one unit of the next smallest layer; 2) every prosodic unit must be dominated by or contained in exactly one unit of the next largest layer; and 3) all phonological content must be parsed into phonological phrases. Languages differ in the number of layers they have and in how these layers are marked, as discussed below.

Because an AM model needs to identify an inventory of contrastive tonal elements in a language, there is no universal AM analysis; rather, each analysis is language-specific, with its own inventory, phonological rules, and representations.

2.3.Prosodic Typology

2.3.1. Word Prosody

Languages differ in how the word is realized prosodically. One can categorize how languages behave based both on the lexical specifications for tone that a language does or does not have, and by the kind of post-lexical marking allowed/required on the word. How researchers classify these languages depends on multiple theoretical viewpoints. In the view proposed by Jun (2005), there are five major categories of word prosodic systems: lexical tone languages, lexical pitch accent languages, stress accent languages, mixed languages, and intonation-only languages.

Taking first one extreme of the typology, we have lexical tone languages, such as Cantonese (Wong et al. 2005), where every syllable in every word is assigned a tone by the lexicon; such languages usually lack word-level post-lexical tones entirely, whether head- or edge-marking. There are also lexical pitch accent languages, such as Standard Japanese (Pierrehumbert & Beckman 1988), where only one syllable of word may be specified for tone. In canonical lexical pitch accent languages, the only reflex of prominence is the tonal event that is specified by the lexicon. Languages may differ in whether a prominent, pitch-accented syllable is required in every word, or whether there is an inventory of different pitch accents that can occur with different lexical items. For instance, in Standard Japanese, there is only one type of pitch accent (a falling tone), and while there can be at most one pitch accent per word, some words have no pitch accent lexically specified. On the other hand, in Serbo-Croatian (Godjevac 2005), there are two possible lexical pitch accents (rising and falling), and every content word must have exactly one. Languages in this group differ in whether they also mark the edges of words with tonal events.

Stress languages, such as English, are similar to lexical pitch accent languages in that there is at most one (and often there *must* be one) prominent syllable in each word. However, rather than being marked with a lexically-determined tonal event, the prominence is realized as stress, which can have any number of acoustic correlates. In English, these include intensity, duration, as well as related segmental processes (e.g. aspiration or tapping). In these languages pitch accents are not specified lexically; instead, they are assigned at a post-lexical level to certain prominent syllables in an utterance as part of the intonational structure, and thus carry sentence-level semantic and pragmatic meanings. Languages vary as to whether the location of a stressed syllable in a word is predictable (as in Arabic) or lexical (as in English). While all stress languages have post-lexical pitch accents, the number of accents in an inventory can vary widely across languages—for instance Egyptian Arabic only has one pitch accent (El Zarka 2013; Helmuth 2006, Chahal & Helmuth 2014), while English has five contrastive pitch accents (Beckman et al. 2005), each with different pragmatic effects.

There are also languages which are considered ‘mixed’ systems (Ladd 1996/2008, Jun 2005, 2014). For instance, Swedish (Bruce 1977, 2005), is usually treated as both a stress language and a lexical pitch accent language, as it has lexically-specified pitch accents, but the syllables to which they attach also show phonetic signs of stress, such as increased duration and intensity. Standard Mainland Mandarin (as spoken in mainland China) is also treated as a mixed system (Shen 1993, Moore 1993, Peng et al. 2005, Jongman et al. 2006), as there is lexical tone specified for nearly every syllable, but stress effects in the form of duration, intensity, in addition to tonal neutralization (partial or total) are still present and perceptible to native speakers.

Lastly, there are languages which lack any lexical specification for intonational or prosodic events, instead relying mostly or wholly on boundary-marking tones to express

phrasing and pragmatic meaning (Jun 2005, 2014). This includes languages like Korean (Jun 1996, 1998, 2000) and West Greenlandic (Arnhold 2014), where regular tonal patterns occur at most word edges.

While these categories have found wide application among prosodic researchers, others, exemplified by Hyman (2006), argue that the distinctions are often muddled. In particular, the designation 'lexical pitch accent' appears to apply to many languages that have actually quite dissimilar prosodic systems, ranging from limited/privative tone systems to a combination of tonal and stress-like behaviors. Hyman thus proposes that only two major determining questions be posed in typologizing prosodic systems: 1) Does pitch play a role in lexical distinctions? If so, the language is tonal; and 2) Does the language require each word to have at least one prominence (obligatory prominence)? If so, the language is stress-accented. Languages of course can have both lexical pitch distinctions and obligatory prominence, and so he uses the relationship between a language's stress and tone systems to further categorize such mixed systems. For instance, if a language's stress assignment is predictable based on lexical tone factors, then the system is described as 'stress dependent on tone.' Conversely, a language where only stressed syllables can bear a lexical tone is one where 'tone is dependent on stress.' Additionally, Hyman identifies a few other parameters that can help to classify prosodic systems, but are not necessarily tied to the classification of a language as being stress-accented or tonal. These parameters are: culminative prominence (at most one prominence per word), prominence subordination (can various prominences be subordinated to one another within a word or phrase?), demarcation (does prominence serve a demarcating effect; i.e. is prominence predictable?), and rhythmicity (are prominences spaced at regular intervals?).

Beyond these characterizations, the word-prosodic systems of broadly tonal languages can be characterized by the domain over which the tonal contrast is realized. In the view exemplified by Jun (2005), this is generally captured within the existing distinction between lexical tone languages, where the syllable is the domain of tone, and lexical pitch accent languages, where the domain of tone is larger than a syllable.

Hyman (2006, 2007) also distinguishes languages by their tonal domains, but due to his avoidance of the term 'pitch accent,' the categories are instead more transparently named 'syllable tone,' 'word tone,' and 'phrasal tone' languages. Word tone languages have a limited number of 'tonal melodies' that can occur over a word. For instance, in Kpelle (Mande family; Welmers 1962:86, via Hyman 2007), both /H.L/ and /M.HL/ sequences can occur over disyllabic words, but the tonal sequence */L.H/ (among many other conceivable patterns) never occurs—this behavior shows that the individual syllables do not actually behave independently in bearing the tonal contrasts. For phrasal tone languages, Hyman (2007) explicitly names Shanghainese as an example, as the contrastive tones are spread across domains that can span lexical word boundaries.

As will become apparent below, Hyman's characterization of Shanghainese is fair; however, there are more complications to the story than he explores in his chapter. While full arguments will be made in the following sections, it is my position that regardless of the terms used or one's view on the term 'lexical pitch accent,' Shanghainese seems to exist as a mixed system, somewhere between a lexical/syllable tone language on one hand and a lexical pitch accent/phrasal tone language on the other.

2.3.2. Phrasal Typology

Languages differ not just by the tonal properties (lexical or not) that the word has, but also in the number of phrasal units that exist above the word. Phrasal units are often recognized by their segmental lengthening effects as well as associated boundary tones, though the exact realizations are always language specific. For example, Cantonese (Wong et al. 2005) and Serbo-Croatian (Godjevac 2005) are argued to have only one level above the word, termed the Intonational Phrase (IP), though they differ in their inventory of IP-final tones. It seems that all languages have at least one level above the word (Jun 2005), but some languages are described as having more, like English with two levels of phrasing above the word (Pierrehumbert 1980, Beckman et al. 2005) and Lekeitio Basque (Elordieta 1997) with three. In general, the first level below the Intonational Phrase ('big' IP) is the intermediate phrase ('little' ip), and the next smallest level is called the Accentual Phrase (AP), which generally dominates the word. As of yet, there are no clear instances of languages with more than three levels of phrasing above the word.

2.4. ToBI Labeling Schema

While autosegmental-metrical analyses are useful in describing the intonational behavior in a language, they do not in and of themselves include a way of transcribing tones or phrasal units. Thus, teams of researchers, as described in Silverman et al. (1992), Beckman & Hirschberg (1994), and Beckman et al. (2005), developed the ToBI (Tones and Break Indices) transcription conventions for 'Mainstream American English,' resulting in MAE_ToBI. In this original ToBI system, an audio recording of an utterance is given with a pitch track, aligned with a waveform of the speech signal. Accompanying the visual representation of the utterance are

four transcription tiers, each with their own function: words, tones, break indices, and miscellaneous. The words tier contains the orthographical representation of each word. The tones tier shows all contrastive tonal events and their alignment, whether head-marking or edge-marking, though predictable tonal effects are not marked. The break index tier shows a numerical value indicating the perceived level of juncture or pause between words (ranging from a juncture smaller than a word boundary to IP-boundary equivalent). Lastly, the miscellaneous tier is used to mark any other noticeable effects in the speech stream or for the transcriber to add comments on the recording. While originally developed for American English, ToBI systems have been developed for multiple languages based on the intonational phonology of the language (e.g., Venditti 2005 for Standard Japanese, Jun 2005 for Korean, Estebas Vilaplana & Prieto 2009 for Spanish; see ToBI systems for more languages in Jun (ed. 2005))—again, just like the AM analyses of intonation for the languages behind the transcription, ToBI models are phonological and thus language specific.

Nevertheless, certain conventions have been adopted in transcribing certain kinds of tonal events. The asterisk '*' is combined with a pitch target, as in H* or L*, to show that the tone is a pitch accent that must align to a prominent syllable. Pitch accents, as referenced above, can contain more than one pitch target; such pitch accents are called 'bitonal' and are shown with a '+' as in L*+H, H*+L, or L+H*. In these cases, the starred tone will always align with the prominent syllable, while the non-starred tone (either leading or trailing) will occur on the immediately preceding (X+) or immediately following (+X) syllable. For IP-boundary tones, '%' is often used, as in L% for a low-target IP-final boundary tone. Such conventions are used in the system proposed in this dissertation, Shanghainese ToBI, or Sh_ToBI.

2.5. Shanghainese Phonetics and Phonology

2.5.1. Shanghainese Citation Tones and their Historical Development

Late Middle Chinese (LMC) is the partially reconstructed ancestor of most currently-spoken Chinese dialects (excluding the Min subfamily). LMC had four lexical tones (*ping,* 'level'; *shang,* 'rising'; *qu*, 'departing/falling'; *ru* 'entering/checked'). Of these, the *ru*/entering tone had the most limited distribution, occurring only in syllables with a coda obstruent, while the rest of the tones could occur on any open or sonorant-final syllable.

In addition to this, LMC contrasted three voicing/phonation types in onset obstruents: voiceless aspirated (only on stops and affricates), voiceless unaspirated, and voiced or breathy. Modal onsets were labeled traditionally as *qing* 'clear,' while voiced/breathy onsets were labeled *zhuo* 'muddy' (Haudricourt 1954, 1961). In the majority of subsequent subfamilies of LMC, this voicing distinction was lost and lead to a full tone split: tones on syllables with breathy onsets became lower, while non-breathy syllables stayed high; breathiness was eventually lost, but its tonal effects stayed, doubling the number of contrastive tones in the resulting varieties (Karlgren 1915-26, Haudricourt 1954, E. Pulleyblank 1978, Yip 1980). Thus, most modern Chinese tone systems are descended from this series of eight tones, which can be classified traditionally as either *yin* tones (originally high register, from a modal onset), or *yang* tones (originally low register, from voiced/breathy onset). Note that while traditional terms like *yingping* (high register level) or *yangqu* (low register departing) are still used to describe tones in today's dialects in current dialectological studies of Chinese, centuries of sound change have obfuscated the relationship between modern contours and these formerly descriptive names.

In contrast to the majority of subsequent Chinese languages, in the Wu subfamily, of which Shanghainese is a member, the LMC phonation contrast was never completely lost, leading to an incomplete tone split. While the *yin*/high and *yang*/low tones are phonetically distinct in pitch and even in the shape of the contour in some Wu languages, both are still accompanied by either voiceless/modal or voiced/breathy consonants, respectively. Subsequently in Shanghainese, the eight surface tones underwent multiple changes and mergers, resulting in the loss of both *shang* 'rising' tones and the one *yangping* or 'low level' tone. These changes resulted in the five citation tones (i.e. the tones as realized in monosyllabic 'citation' contexts) of Shanghainese. The table below shows these five citation tones with Chao tone numbers as given by Xu & Tang (1988), along with their traditional numbers and categories, though others have transcribed the same citation tones differently (Zhu 1995, Chen & Gussenhoven 2015).

Basic Contour	**-Glottal Rhyme Open**		**+Glottal Rhyme[1] Checked**	
	-Breathy Onset H Register	**+Breathy Onset[2] L Register**	**-Breathy Onset H Register**	**+Breathy Onset L Register**
/HL/	T1 *Yinping* [53] 爸 *pà* 'dad'			
/LH/	T2 *Yinqu* [34] 摆 *pa* 'to put'	T3 *Yangqu* [23] 败 *ba* 'to fail'	T4 *Yinru* [<u>55</u>] 八 *paʔ* 'eight'	T5 *Yangru* [<u>12</u>] 白 *baʔ* 'white'

Table 1. The five citation tones of Shanghainese, showing basic contour, traditional numbering (e.g. Tn), tonal category, and phonotactic environment. Tones are transcribed using Chao tone numbers (Chao 1948). Underlined tone numbers indicate shortened duration due to glottal coda. Tonal values taken from Xu et al. 1986.

[1] Note that all coda obstruents in Late Middle Chinese merged to /ʔ/ in Wu dialects.
[2] Note that breathy obstruents will be written with their fully voiced equivalents.

In traditional Chinese dialectology (exemplified by Xu et al. 1986), citation tones are taken at face value and considered basic; thus any tonal phonological process will operate over the tones as separate categories. This stands in contrast to generative analyses of the tonal inventory (Zee & Maddieson 1980, Selkirk & Shen 1990, Duanmu 1993, M. Chen 2000,), which generally seek to minimize the number of phonemic contrasts, deriving the surface 'citation' tones via rules or constraints from some abstract tonal representation. Under this kind of analysis, the Shanghainese tonal inventory is reducible to two tonal contours—a fall /HL/, and a rise /LH/[3]—and the five citation tones thus arise from the combined effects of the tone's segmental environment: breathy onsets lower tones; glottal codas flatten and/or shorten them. A static phonotactic ban on /HL/ occurring with either breathy onsets or glottal codas must also be assumed. Many phonological studies, such as the ones listed above, take this view of the tonal system, but the phonetic reality of the relationship between phonation and pitch is far more complicated, as shown in the following section. Nevertheless, the usefulness of such an abstract representation in describing tonal behavior will also be acknowledged in the model presented in this dissertation.

It is important to note that there is some debate about the nature of the phonation contrast as it applies to sonorant onsets, such as /m/ and /l/, which can occur with high or low register tones. While some scholars have thus transcribed these onsets as contrastive (for example: /ʔm/ vs. /ɦm/) (Xu & Tang 1988), others (Zee & Xu 2016) accuse such analyses of forcing a transcription that idealizes the relationship of phonation and tone rather than accurately describing the realization of the segment(s) in question. As will be described in greater detail in

[3] Though the high glottal tone, T4, is transcribed as level, its behavior in sandhi contexts generally leads generative analyses to treat it as a rising /LH/. This will be covered in greater detail in Section 2.5.4.1.

the following section, there is in fact a phonation distinction on sonorants, though speakers vary widely in their production of the breathiness contrast in general, with sonorants presenting the largest differences between speakers.

2.5.2. Introduction to Shanghainese Breathy Voice and Ongoing Language Change

The phonation contrast in Wu dialects has been the subject of linguistic discussion and research for nearly a century, with the first modern accounts given by Liu (1924) and Chao (1928); in the latter of these, the breathy voice is qualitatively described as a "clear sound with muddy flow." This has been taken to describe the general realization of non-modal phonation in Shanghainese stops as voiceless during closure followed by breathy/muddy phonation on the following rhyme, which has been phonetically confirmed by more recent studies (Cao & Maddieson 1992, Z. Chen 2010). However, it is also noted in multiple studies that breathy obstruents are realized as fully voiced in compound words between vowels (Cao & Maddieson 1992, Z. Chen 2010, Y. Chen 2011, Gao & Hallé 2017) as part of a sandhi process.

It has been termed alternately 'breathy' or 'murmur' by Sherard (1972), Ramsey (1989), Cao & Maddieson (1992), and Zhu (1995), while Ladefoged & Maddieson (1996), Y. Chen (2011), and Gao et al. (2011) specifically characterize it as 'slack,' due to the apparent greater glottal aperture during production, distinguishing it from a more canonically breathy phonation contrast, like the one in Gujarati (Khan 2012). Most recently, Tian & Kuang (2019) use the term 'whisper' to describe the phonation type, due to the heavy influence of acoustic noise in its production by many of their speakers. In the remainder of this dissertation, 'breathy voice' or simply 'voiced' will be used to refer to this contrast. While some impressionistic studies have

described this breathiness as a property of the whole syllable (Ramsey 1989), others generally agree that the phonetic data show that breathiness, when present, occurs on the first half of the nucleus following the breathy consonant (Cao & Maddieson 1992, Ren 1992, Zhu 1995, Tian & Kuang 2016). Despite these general behaviors, most phonetic studies show that the realization and strength of breathy phonation varies widely based on manner of articulation, prosodic position, and speaker age.

Shanghainese has undergone rapid language change in the last century, no doubt due in part to the rise of Standard Mandarin in formal settings and the dramatic influx of domestic immigrants, all bringing their own languages and/or dialects. Because of this demographic shift, dialectologists have distinguished three brackets of speakers based on their year of birth: 1) Old Shanghainese: speakers born before 1940; 2) Middle Shanghainese: speakers born approximately 1940-1970; and 3) New Shanghainese: speakers born approximately 1970-1990 (Xu & Tang 1988, Qian 2003). However, as we will see, speakers born in the 1970s have been grouped with both 'younger' and 'older' speakers in the different studies outlined below. The generational changes affect multiple parts of the sound system, leading to fairly large differences among the generations. Note that in this dissertation, all speakers analyzed were born between 1937 and 1975 (median year: 1965), so the vast majority are speakers of the Middle variety.[4]

In regards to contrastive breathiness, the literature shows a trend of younger speakers losing the phonation contrast, tending towards a purely tonal contrast, essentially going the way of the rest of the Chinese subfamilies in nearing completion of a full tone split. On the other

[4] See Chapter 3, Methodology, for more information on ensuring that all speakers had comparable dialects.

hand, Middle Shanghainese speakers generally maintain the phonation contrast, though pitch remains an invaluable cue for production and perception of the distinction.

2.5.3. Phonetic Studies of Shanghainese Breathy Voice

Articulatorily, breathiness is often characterized as having less glottal constriction and thus greater airflow during phonation. Studies using electro-glottographs (EGG), which measure the electrical impedance across a speaker's vocal folds over time during speech, will either report Open or Closed Quotients (OQ and CQ), which estimate the ratio of time the vocal folds spent apart versus closed together in one cycle. As opposite measures, higher OQ and lower CQ (i.e., more time spent with folds apart) correlate with breathiness in many languages. In addition to EGG, breathiness can also be measured by oral airflow and intra-oral air pressure, which can be measured using flow masks and pressure transducer, respectively. Results from such measurements are usually reported in a ratio of airflow to air pressure, with higher values (i.e. more relative airflow to air pressure) being associated with breathy phonation.

In terms of acoustics, breathy phonation is often characterized by higher values of spectral measures (e.g. H1-H2, H1-A1, H1-A2), which compare amplitude of the first harmonic of the sound wave (H1) to the amplitude of either higher harmonics (H2), or harmonics closest to formants (A1, A2). Higher spectral measures thus show a higher relative amplitude of lower frequencies, or spectral tilt, which is correlated with contrastive breathy voice in many languages, including Jalapa Mazatec (Garellek & Keating 2011), Gujarati (Khan 2012), White Hmong (Esposito 2012), and Southern Yi (Kuang & Keating 2014).

While spectral measures are usually the best indicator of breathiness across languages (Tian & Kuang 2016), other acoustic measures can be used to determine the amplitude or relative amount of noise or aperiodicity in the speech signal. Most common among these measures is Harmonic-to-Noise Ratio (HNR, de Krom 1993), which measures the relative energies of periodic vs aperiodic sound in specifically selected frequency bands (e.g. 0-500Hz, 500-1500Hz, etc.). Lower HNR values indicate more relative noise, and thus a less periodic, less modal phonation type. Another of these aperiodicity measures is Cepstral Peak Prominence (CPP, Hillenbrand et al. 1994), which is assumed to also reflect harmonic-to-noise ratio, though it is calculated differently, namely, by identifying the highest amplitude frequency in a cepstrum of the sound wave. The amplitude of this peak is compared against the background noise, giving a relative strength of cepstral peak. Again, lower values are correlated with breathier phonation.

Multiple production and perceptual studies have investigated the role of such cues in voicing contrast in Shanghainese. In earlier studies (Ren & Mattingly 1989, Cao & Maddieson 1992, Ren 1992), focus was almost entirely on the spectral and articulatory properties of the voicing distinction in stops, with few other variables. These studies all show the significant correlation of H1-H2 and H1-A1 with the phonation contrast in production (Ren & Mattingly 1989, Cao & Maddieson 1992, Ren 1992) and perception (Ren 1992) of monosyllables. In compound-medial sandhi position, again, all three studies found that the usefulness of these measures lessened (with the exception of H1-A1 in Ren 1992). In terms of articulatory measures, Cao & Maddieson (1992) find a higher airflow/air pressure ratio for breathy bilabial stops, while the other two, Ren & Mattingly (1989) and Ren (1992), find that open quotient (OQ) of the glottis is higher for voiced stops. Ren (1992) also finds a significant effect of f0 on the perception of stops, but argues against pitch as the primary cue of the voicing contrast, due to the

fact that H1-H2 and H1-A1 are higher in both voiced and voiceless aspirated stops, which co-occur with high and low tones respectively. Only Ren (1992) gives an approximate range of birth years for his speakers (~1957-1965); however, it is assumed that these earlier studies overall are most relevant to the distinction as it is made by Middle Shanghainese speakers.

In later studies, many more variables were considered in the production and perception of the breathiness contrast. Y. Chen (2011) specifically investigated the production of the voicing contrast in compound-medial stops in older speakers (born 1935-1950), showing that spectral measures generally became less useful. On the other hand, she finds that full voicing of stops in the same position is common, but influenced by the tone of the preceding syllable. Gao's (2015) dissertation and related publications (Gao & Hallé 2013a, Gao & Hallé 2013b, Gao & Hallé 2015, Gao 2016, Gao & Hallé 2017, Gao et al. 2020) constitute perhaps the most comprehensive study of the voicing contrast in Shanghainese. In addition to confirming previous production findings that spectral measures and f0 are significant contributors to the contrast, she further shows that older men (born 1938-1956) have the strongest spectral cues (Gao & Hallé 2013b, Gao 2016, Gao & Hallé 2017) which are present in all manners (Gao 2016), and that losses of spectral cue strength are seen in both female speakers and younger speakers. Thus, young women make almost no phonation contrast in her studies; however, there is a trade-off relationship, where some younger speakers (particularly men) still used spectral measures if they had narrow pitch ranges (Gao & Hallé 2013b). Nasals also have little difference in spectral measures in her research, though they trend in the right direction (Gao & Hallé 2017).

Beyond spectral measures, she also finds that older speakers make use of noise measures, such as HNR and CPP, to distinguish the phonation contrast as well, though its affects are strongest following stops, with fricatives next, and sonorants showing almost no difference in

noise measures (Gao 2016, Gao & Hallé 2017). Additionally, these speakers' noise and spectral cues are weakened in compound-medial position. As with the spectral measures, she finds that these noise cues are basically absent in production for younger speakers (born 1985-1996) (Gao 2016, Gao & Hallé 2017). Interestingly, in a perception experiment, younger speakers were still found to make use of spectral and noise measures in perceiving the distinction, despite not using these cues in production (Gao 2015, Gao et al. 2020).

In terms of articulatory measures, Gao & Hallé (2013b), Gao (2016), and Gao & Hallé (2017) all find that OQ is significantly higher in production of voiced consonants for older men (born 1938-1956), though there was one exception of a young male using OQ in Gao & Hallé (2017). No effect of manner was reported in Gao & Hallé (2013b) and Gao (2016); however, in Gao & Hallé (2017), OQ is reported to be significant only in the production of stops and fricatives, not sonorants.

Beyond these regular measures for breathiness, Gao & Hallé also investigated the role of phonetic cues associated with voicing, such as consonant duration, voice-ratio (v-ratio, comparing the duration of voicing to the duration of voicelessness during a consonant), and voice onset time (VOT, the time between consonant release and onset of voicing) for stops. In the case of duration, Gao & Hallé (2013a) found that stop and fricative durations were a strong cue in addition to the primary f0 cues in perception of voiced consonants for younger speakers (born 1979-1995). Gao & Hallé (2017) investigated the most of these voicing measures in production, finding that older speakers (born 1938-1956) produce the breathy/voiced stops with longer VOT. Despite confirming previous studies that breathy stops are voiceless word-initially and voiced word-medially, they also found that breathy fricatives are often fully voiced even in initial position—this trend was stronger for younger speakers, women, and the labial fricative /v/.

Two studies by Tian & Kuang (2016, 2019) also investigate many of these same cues, and while they generally find the same trends across age groups and prosodic positions, the differences between the breathy and modal consonants are lessened across manner when compared to others' results. Tian & Kuang (2016) compares the production of the phonation contrast of older (born 1933-1971) and younger (born 1984-1999) speakers, investigating spectral measures (H1-H2, H1-A1, H1-A2) and noise measures (HNR, CPP), and finds that H1-A1 and H1-A2 are only reliable for older male speakers, contradicting earlier research on H1-H2's role in the contrast. On the other hand, the noise measures, especially CPP, can reliably identify the contrast in older speakers of both genders, in any manner. Younger speakers showed little breathiness via these measures, though their noise measures trended in the expected direction. Tian & Kuang (2019) specifically turn to 'older speakers' (born before 1980) and their production, with perhaps the largest number of participants of any study (n=52); again confirming the results of their 2016 study—noise measures are better than spectral measures, with no effects of gender or manner, including sonorants. Despite the overall primacy of noise measures in the aggregated data, they note significant individual variation, with a large minority (~23%) of speakers using more spectral cues in their production. Both studies confirmed that older speakers also have a lower closed quotient (CQ) following voiced stops, indicating a more open glottis and thus breathier voice.

The last study presented here is Zhang & Yan (2018), which unlike Tian & Kuang (2019), focused exclusively on young speakers—in their production task the average birth year was 1990, while in the perception task, it was 1994. The results from the production experiment show differences in how the contrast is cued both across manners and in different positions. In initial contexts, f0 is by far the strongest cue. While spectral and noise measures trended in the

expected directions for all manners in this position, including sonorants, only CPP was found to have a significant effect, and then only within fricatives. They argue that the CPP results during monosyllabic fricatives and sonorants show that the production of at least some consonants play a role in the phonation contrast. They also propose that the f0 differences are in part due to the consonantal contrast, as the difference in f0 following voiced and voiceless consonants diminishes across the vowel, indicating an effect originating from the beginning of the syllable. In medial contexts, the f0 cue was neutralized, and most phonation cues, both spectral and noise, were lost. Again, fricatives maintain a CPP difference in the contrast. Despite the loss of the phonation and tonal cues in medial contexts, stops and fricatives became voiced and shorter, while sonorants lost any trace of distinction.

In Zhang and Yan's perception study, similar results were found with stimuli spliced from real speech, where both consonantal and pitch effects were found to contribute to the categorization of voicing/breathiness, with differences along both position and manner. They present two statistical analyses to explain the data: a more conservative Classification and Regression Tree (CART) and a more inclusive logistic regression model. In both cases, for monosyllables, consonantal cues (e.g. duration, voicing) and f0 contributed significantly to the identification of stop and fricative voicing, with f0 able to override any other conflicting cues. The logistic regression additionally found that phonation cues are also used in monosyllabic stops and fricatives. In disyllables, the logistic regression found that f0 lost its usefulness in stops, but not fricatives, and that consonantal cues still contributed to stop identification.

In summary, the Shanghainese phonation/voicing contrast is made through a complex array of cues (spectral, noise, voicing, duration, and pitch), whose relative strengths are subject to various influences including age, gender, prosodic position, and manner. In initial or

monosyllabic contexts, the contrast is mostly pitch and phonation based, while in medial contexts, the contrast becomes mostly one of voicing. While younger speakers are losing many of the phonation cues (spectral, noise, OQ, airflow) associated with the distinction, they are often shown to maintain a voicing distinction in medial position, and even with initial fricatives, so the consonantal contrast has not been completely lost; sonorants on the other hand have lost any trace of this contrast. Despite this loss of breathiness in production, younger speakers are still aware of its usefulness as a cue in perception, no doubt due to their continued exposure to older speakers. For older speakers, the phonation contrast is more robust, with male speakers having the strongest breathiness cues of any population. Again, the same trend of stops and fricatives having the clearest phonation cues arises in older speakers, though unlike younger speakers, sonorants were found to actually have phonation differences, even if weakened in comparison to other manners. This means that while a tone split has truly begun in New Shanghainese (especially with sonorant-onset syllables), Middle Shanghainese speakers still have high and low tones in complementary distribution. Unfortunately at this time, there are no perception experiments carried out specifically with older speakers to test the relative strength of these phonation and voicing cues to f0 cues in categorization. While studies working with younger speakers can safely conclude that pitch is the primary cue to the contrast (at least in monosyllables), such a claim cannot be made for older speakers yet.

2.5.4. Shanghainese Tone Sandhis

Many Chinese languages are famous for their tone sandhi systems; M. Chen (2000) gives the most basic definition of tone sandhi as “contextually determined tonal alternation” (p. 23). For most Chinese tone sandhi processes, the conditioning contexts are either nearby tones or

prosodic position. However, beyond this simple definition, researchers have long noted the asymmetry of sandhi patterns in Chinese languages, categorizing them as either left- or right-dominant (Yue-Hashimoto 1987, Chan 1995, M. Chen 2000, among others). These terms are used to designate which syllable in a given adjacent pair will undergo some sort of tonal alternation and which will maintain its underlying or citation tone in some sense. For example, the famous Standard Mandarin third tone sandhi is an example of right-dominant sandhi (henceforth RD sandhi), as the right syllable in a sequence of two low 'third' tones remains low, while the left syllable undergoes sandhi to become a rising 'second' tone. RD sandhis have a tendency for paradigmatic replacement of tones, with one citation tone replacing another in specific tonal contexts (Zhang 2007). On the other hand, in left-dominant sandhis (henceforth LD sandhi), the leftmost member of a domain maintains its tone, while all others lose theirs. Again, the most familiar example comes from Standard Mainland Mandarin in the form of syllables with the fifth or 'light' tone, whose pitch is determined by the tone of the immediately previous (or left) syllable. However, unlike most RD sandhis, this 'light' pitch is often not the same as any citation tone in the inventory of the language (Zhang 2007). Many other LD sandhi patterns, such as those found in Wu dialects, also involve spreading the left syllable's tone over the domain, with non-domain-initial syllable(s) losing most traces of their citation tone(s).

As is evidenced by Mandarin, many Chinese languages have both LD and RD sandhi processes—Shanghainese is another of these. In this section, I will review the existing literature on the sandhi processes in Shanghainese, starting first with its more well-studied LD sandhi before outlining the behavior of its RD sandhi. In general, the literature finds the spreading LD sandhi is a consistent and productive phonological process; however, generational effects come in to play, with significant differences between young and old speakers as to the phonetic

realization of LD sandhi in production. On the other hand, the characterization of the less-studied RD sandhi has been complicated by modern research, showing that it perhaps is best characterized as a kind of tonal reduction. This section will focus primarily on the basics of the sandhi processes and their local phonetic implementation, followed by a review of previous prosodic literature concerning how sandhi domains are determined in various contexts, and how they are produced under focus conditions.

2.5.4.1. Shanghainese Left-Dominant Sandhi

One of the most famous features of Shanghainese tone is its left-dominant spreading sandhi pattern (LD sandhi), also known as 'broad-use' sandhi by Xu & Tang (1988). This sandhi pattern has been discussed fairly extensively in the literature (Sherard 1972, Zee & Maddieson 1980, Shen 1981, Xu et. al 1981, Jin 1986, Xu & Tang 1988, Toda 1990, Zhu 1995, Zhu 2006, and others), and is generally described as a process where the tonal targets of the first syllable in some multisyllabic sandhi domain are spread across the domain, associating each tonal target to a syllable in a one-to-one, left-to-right fashion. Since all tones are either rising or falling, there are only two tone targets that are spreadable; thus, any domains over two syllables will have some 'default' tone inserted on third and later syllables (see T5 behavior below for an exception). Regardless, if a non-initial syllable receives a tonal target from the initial syllable or a default tone, the result is the phonological deletion of all tonal information on non-initial syllables. Sandhi domains generally consist of a word (often a di- or tri-syllabic compound, as is often the case in Chinese languages) and any light particles following said word, though domains up to five syllables are observed.

In terms of motivating the sandhi system broadly, Duanmu (1993) introduces the idea that Shanghainese sandhi is a realization of the stress difference between initial and non-initial syllables (developed further in Duanmu 1995, 1997); other researchers have also made this assumption, whether working in a more generative/phonological framework (Zee & Maddieson 1980) or a more traditional/phonetic one (Ren 1992, Zhu 1995, Gao & Hallé 2013a). The idea is that Shanghainese has fixed, word-initial stress, and that non-stressed syllables are not permitted to realize their tone. Duanmu (1995) then explains the spreading as the result of a constraint in the language that all syllables have some assigned tone. While Duanmu admits that stress usually implies the existence of specific phonetic cues (e.g. duration, intensity, etc.), which may not be perceptually salient in Shanghainese, some sort of phonological privilege or prominence is useful to the characterization of LD sandhis; the analysis of Shanghainese presented below is no different.

The following two tables show, in schematized form, the basic pattern of the LD sandhi as understood by traditional/surface-oriented accounts (Table 1, ex. Xu et al. 1986) and abstract generative accounts (Table 2, ex. Duanmu 1993). In both tables, the left column shows the citation tone of the domain-initial syllable, while each subsequent column lists the tones for the entire sandhi domain given a number of syllables. As stated before, since non-domain-initial syllables lose their underlying tones, there is no need to take their pre-sandhi tones into account.

Initial σ Tone	**1 σ**	**2 σ**	**3 σ**	**4 σ**	**5 σ**
T1 *High*	[53]	[55 21]	[55 33 21]	[55 33 33 21]	[55 33 33 33 21]
T2 *High*	[34]	[33 44]	[33 44 21]	[33 44 33 21]	[33 44 33 33 21]
T3 *Low*	[23]	[22 44]	[22 44 21]	[22 44 33 21]	[22 44 33 33 21]
T4 *High, glottal*	[<u>44</u>]	[<u>33</u> 44]	[<u>33</u> 44 21]	[<u>33</u> 44 33 21]	[<u>33</u> 44 33 33 21]
T5 *Low, glottal*	[<u>12</u>]	[<u>11</u> 33]	[<u>11</u> 33 21]	[<u>11</u> 33 22 21] [<u>11</u> 11 11 12]	[<u>11</u> 33 22 22 11]

Table 2. Left-dominant sandhi patterns of Shanghainese under traditional analysis, including register and glottal information for each of the citation tones. Underlined Chao numbers indicate a shortened, glottal tone. Modified from Xu & Tang (1988), pg. 24.

Initial σ Tone	**1 σ**	**2 σ**	**3 σ**	**4 σ**	**5 σ**
T1: /HL/ *High*	[53]	H L	H L L	H L L L	H L L L L
T2: /LH/ *High*	[34]	L H	L H L	L H L L	L H L L L
T3: /LH/ *Low*	[23]	L H	L H L	L H L L	L H L L L
T4: /LH/ *High, glottal*	[<u>44</u>]	L H	L H L	L H L L	L H L L L
T5: /LH/ *Low, glottal*	[<u>12</u>]	L H	L H L	L H L L L L L H	L H L L L

Table 3. Left-dominant sandhi patterns of Shanghainese under generative analysis, including register and glottal information as well as underlying contour for each of the citation tones. Underlined Chao numbers indicate a shortened, glottal tone. Modified from Xu & Tang (1988), pg. 24.

The most obvious difference here is the amount of phonetic nuance. In the generative analyses, the tones and thus the sandhi process are stripped down to their most basic components, while in the traditional analysis, we gain more insight into the actual realization of the pitch targets, though the numerical transcriptions can still be impressionistic and/or idealized. Regardless, in the numerical transcriptions, we can see for instance that the phonation/pitch

distinction of the initial syllable is apparently kept, with T2 and T3 differing in their starting points, even in multisyllabic domains. While under an abstract generative analysis, this behavior may be derivable from onset voicing, there is problem for the inserted default tones. In the generative account, only abstract L tones are used, while in the traditional one, we see a range of values—[33], [22], [11], [21], [12]—depending both on position of the syllable in the domain or the tone of the domain-initial syllable. While potentially interpretable as all Low targets, the variation in pitch indicate that there is more going on in these default tones than had been investigated in purely phonological studies. Below, phonetic work on the realization of default low tones is presented, showing again large differences based on the age of the speaker.

Another peculiarity here is that T5, the low rising glottal tone, apparently does not always follow the regular pattern in four syllable domains as described by Xu & Tang (1988). Instead, quadrisyllabic T5-initial domains assign a default low target to all non-final syllables in the domain while assigning a H target to the final syllable—rather than tonal spreading, this alternation has been dubbed tonal displacement (Zhu 1995; Zhang & Meng 2016). Some may argue that right edge is somehow favored or prominent in these domains; however, the overall tone on such a domain is still determined by the domain-initial syllable, making it an example of LD sandhi. Researchers differ in their exact characterization of this variation, or whether variation even exists. Earlier studies, such as Xu & Tang (1988), simply state that the process is fully variable, with certain words and phrases perhaps preferring one pattern over the other. Zhu (1995) has perhaps the most restrictive description of tonal displacement, proposing that this pattern can only and must occur when the second syllable has a non-glottal rhyme, though no sandhi domain length restrictions are mentioned. In Zhang & Meng's (2016) paper, LD sandhi production in disyllabic domains was phonetically analyzed, with the conclusion that

displacement occurs even in disyllables, with T5-initial domains consisting of a low initial syllable and a second syllable with a low rise, [12], statistically equivalent to monosyllabic T5 realization; again, this result shows that tonal displacement can apply in domains shorter than four syllables. In a different set of studies, Takahashi (2013, 2015) investigates T5 left-dominant sandhi specifically in quadrisyllabic domains, and reaches the conclusion that the choice of spreading versus displacement is due to internal morphological structure of the phonological phrase. In his study, he finds that tonal displacement is essentially the *only* option for T5-initial quadrisyllables that do not have a [2+2] internal structure (i.e. consisting of two disyllabic elements; here, nouns or nouns with a sentence-final particle). For these [2+2] quadrisyllables, he finds instead that the normative spreading pattern is preferred (59% of tokens) to tonal displacement (23%), and that some simply split into two separate sandhi domains (16%). This story paints a less variable picture than that given by Xu & Tang (1988), but more variable than the conditions proposed by Zhu (1995). This issue of T5 variability complicates the generative analyses that posit only two underlying tones in Shanghainese, as we would expect T5's behavior to be completely in line with the other rising tones (T2-4). It is for this reason that I will treat T5 as a separate third phonemic category in this dissertation, different from falling T1 and rising/spreading T2-4. It should also be noted that in the current study, almost all speakers used the tonal displacement pattern in all T5-initial domains, regardless of length, and while two speakers did use the standard spreading with some lexical items, no speakers varied their production of these specific words; further discussion of my speakers' T5 behavior will follow in Section 4.2.

As is always the case, in the phonetic realm, the picture of LD sandhi is not nearly as clear cut as the tables above would make it seem. As already mentioned, of particular interest is the phonetic realization of domain internal ‘default lows’ that are inserted on every syllable after the second one in a sandhi domain. Impressionistic accounts, like Xu & Tang (1988), describe the default low as a non-prominent mid tone (usually [22] or [33] in Chao numbers) that falls fully on the final syllable to a true low target (usually [21] or [11]). However, more modern experimental studies have shown different results that seem dependent on the generation of the speaker. Y. Chen (2008) investigates quadrisyllabic domains and finds that, with older speakers (born 1935-1950), the most common strategy is interpolation from the previous H target to a low point at the end of the sandhi domain. This result suggests that there is not a separate phonological target that appears on the third syllable of a domain; rather, there is a low target aligned to the end of the domain. She also notes that the final, inserted L targets are weaker compared to the lexical L targets associated with initial syllables, comparing the default L to the toneless ‘light’ syllables of Mandarin, echoing the idea of phonological prominence forwarded by Duanmu (1993) and Zhu (1995).

In contrast to these results for older speakers, Takahashi (2013) tests the production of ‘default lows’ in younger speakers (born 1983-1988) and comes to the conclusion that they indeed do have a phonological L-tone insertion on the third syllable of sandhi domains. Y. Chen (2008) notes that her youngest speakers show some variation along these lines, looking sometimes like the younger speakers in Takahashi’s study. Y. Chen and Takahashi both take this to be a sign of sound change with later influence of younger speakers on older speakers.

Another phonetic fact that is not captured by either of the tables above has to do with the amount of influence of the initial syllable’s citation tone one the realization of the whole domain.

Chen (2008), working with older speakers (born 1935-1950), showed that the original citation tone of the initial syllable can still be statistically detected via f0 on the second syllable and into the third syllable of a domain, though its effect dissipates before the fourth. To give an example, this means that the syllable after a domain-initial high rise (T2) or low rise (T3) consistently differ, regardless of the fact that they are predicted to be the same 'high' tone target ([44] in traditional accounts, Table 2 above; /H/ in generative analyses, Table 3).

There is also disagreement over the extent to which the second syllable of a domain can influence the tonal realization, post sandhi application. Zhang & Meng (2016), who worked with disyllables produced by younger speakers (average birth year: 1990), actually found that while the effect of the first syllable's tone on the second syllable's pitch is significant, there is also a significant tonal input from the second syllable's citation tone; this is especially true when the initial syllable was a nonce morpheme. However, they stop short of fully rebutting earlier claims of full neutralization, as their speakers were primed with the citation tones for both syllables before producing the target disyllable. In a smaller production study with similarly young speakers Ling & Liang (2018) find no such effect. In perception, Zhang & Yan (2018) confirmed the effect of non-initial citation tones only for medial fricatives for younger speakers. It is unclear if the second syllable's citation tones affects the production of pitch contours in older speakers' production or perception, as no study has focused on this. Kuang et al. (2018) find similar results in closely-related Northern Wu dialects (namely Shaoxing, Zhuji, and Xinzhuang, all spoken within a ~225km radius of urban Shanghai), where both initial and non-initial tones can contribute to sandhi domain pitch realization both phonetically and phonologically. These more recent studies acknowledge the fact that LD sandhi itself can be a gradient process, with more or less influence coming from non-initial citation tones, though it remains unclear to what

extent speakers in older age ranges lack complete neutralization of the non-initial tones in a sandhi domain.

Multiple studies have tested the psychological reality of LD sandhi, though in different ways, and with interesting results. Zhang & Meng's (2016) study tested the productivity and phonetic realization of Shanghainese LD sandhi, asking participants (younger speakers, born on average ~1990) to apply the sandhi to monosyllables presented auditorily, whether real or nonce. They find that LD sandhi is productive with nonce words, showing that speakers have knowledge of the process and are able to apply it to novel forms. This suggests a more syllable-tone than phrasal tone type system. On the other hand, Kuang & Tian (2019) conducted a speeded production task, where both older (born 1941-1981) and younger (born 1985-2002) speakers were shown a mono- or di-syllabic word in Chinese characters and asked to name it out loud. The response time was measured and taken as an indication of the amount of time needed to access the lexical item. Counter to the expectations following from Zhang & Meng (2016), Kuang & Tian's speakers, regardless of age, took longer to name monosyllabic words than disyllabic words. For monosyllables that had low frequency in word-initial position, the reaction time was even greater. Kuang & Tian (2019) argue that these results show that Shanghainese speakers store *both* complete sandhi forms and monosyllabic representations in their lexicon, and that the sandhi forms are more readily accessed; this is likely due to the fact that most words in Shanghainese are compounds, and thus most words are heard more frequently in sandhi form than in citation form. This effect is further confirmed by Yan, Chien, & Zhang (2020), who conducted lexical decision tasks with disyllabic targets. These disyllables were preceded by either an 'underlying' prime (i.e. syllable with the same citation tone as the first syllable in the target disyllable) or a 'surface' prime (i.e. a syllable taken from a disyllabic form with the same

initial tone as the target disyllable). Two experiments were run, with generally younger speakers (born 1982-2002) who used Shanghainese less frequently (self-reported <50% of the time) and with generally older speakers (born 1957-1980) who use Shanghainese more frequently (>50% or the time). It was found that only surface primes had a facilitatory effect, and only for less frequent users of Shanghainese. This indicates that disyllables are most likely stored whole, with sandhi already applied. The lack of any priming effects in more frequent users is attributed to lexical competition overriding phonological priming due to their potentially larger lexicon than less frequent users. Nevertheless, the lack of underlying priming effects in more frequent users is surprising given the productivity of the sandhi in general—again, this shows that the monosyllabic citation forms of tones are not always directly linked with the sandhi forms that seem to be stored separately.

To summarize, Shanghainese left-dominant sandhi is characterized by the tone of a prominent, domain-initial syllable spreading over the first two syllables of said domain. In domains longer than two syllables, older speakers will interpolate from the previous tonal target to a weak low tone at the end of the domain, whereas younger speakers will insert a default low on the third syllable, causing a low plateau to the end of the domain. T5, the low register glottal tone, has apparently variable behavior in LD sandhi, sometimes patterning with the other tones, and sometimes undergoing tonal displacement, where its citation contour moves to the final syllable of the domain, leaving a low plateau before it. However, the nature of the variation is contested and poorly understood. In terms of phonetic studies, domain-initial citation tones can influence the tonal shape of the entire domain, with some researchers finding influence from the citation tones of domain-second syllables. Psychologically, LD sandhi is a productive process

that can be applied in novel situations, but other experiments suggest that despite this productivity, the vast majority of words in Shanghainese are actually stored whole, with ‘sandhi’ already applied. Thus it appears that both ‘underlying/citation’ and ‘surface/sandhi’ representations are available concurrently to speakers of Shanghainese, making it hard to fit LD sandhi into a straightforward generative account where stored underlying forms are manipulated by the grammar to produce non-stored surface forms.

2.5.4.2. Shanghainese Right-Dominant Sandhi

Although much less discussed in the literature, Shanghainese also has a secondary, right-dominant (RD) sandhi process, also called ‘narrow-use’ sandhi by Xu & Tang (1988). Traditionally, this is described as the last syllable sandhi domain neutralizing to some level mid tone while the following domain-initial syllable remains unchanged in tonal realization. The following table, modified from Xu & Tang (1988) pg. 25, shows each citation tone with its corresponding predicted neutralized tone:

<table>
<tr><th></th><th>T1
[53]</th><th>T2
[34]</th><th>T3
[23]</th><th>T4
[44]</th><th>T5
[12]</th></tr>
<tr><td>Monosyllabic SD
(x) (σ …)</td><td colspan="2" rowspan="2">[44]</td><td rowspan="2">[33]</td><td rowspan="2">[44]</td><td rowspan="2">[22]</td></tr>
<tr><td>1σ SD following 1σ SD
(σ) (x) (σ …)</td></tr>
<tr><td>Last σ in multi-σ SD
(… σ x) (σ …)</td><td colspan="3" rowspan="2">[33]</td><td colspan="2" rowspan="2">[33]</td></tr>
<tr><td>1σ SD after multi-σ SD
(… σ σ) (x) (σ …)</td></tr>
</table>

Table 4. Shanghainese right-dominant sandhi patterns using Chao tone numbers. Underlined tones are shortened/checked. Left-hand column shows target syllable *x* in varying prosodic contexts; columns represent different citation tones for target *x*, with citation forms in top row. SD stands for sandhi domain. Modified from Xu & Tang (1988), pg. 25.

In this process, we see four different configurations which can lead to tonal neutralization. The first two rows show a monosyllabic sandhi domain (SD) undergoing the same neutralization whether it occurs in sentence-initial position (as in the first row) or in sentence-medial position following another monosyllabic SD (second row). In these cases, we can see that there is not a full neutralization of tone—a distinction between breathy (T3, T5) and non-breathy (T1, T2, T4) onset syllables is maintained. Additionally, a durational cue remains to distinguish glottal (T4, T5) and non-glottal (T1-3) rhymes as well. In the next two rows, we can see a more drastic neutralization pattern that can affect both the last syllable in a multi-syllabic SD (third row) and monosyllabic SDs that are preceded by multi-syllabic SDs. Again, duration cues help to differentiate non-glottal T1-3 from glottal T4 and T5. Note that in situations like the third row, where the neutralized syllable is at the end of a longer, multi-syllabic SD, LD sandhi spreading appears to occur "first," with any neutralization applying after tonal targets have been reassigned within the SD. While Rows 1, 2, and 4 all involve monosyllables, the difference between them lies in what precedes them (i.e. nothing, another monosyllable, or a multi-syllabic SD, respectively). Xu & Tang (1988) also note that beyond the tonal neutralizations, syllables undergoing RD sandhi are followed by a "small pause."

In Xu et al.'s in-depth description of Shanghainese, they make sure to stress the variability in the application of this sandhi pattern. It applies most often on monosyllabic domains, which are, by and large, verbs. Indeed, there seem to be many lexicalized verb-object disyllables (a common features of Chinese languages) where the verb undergoes right-dominant (RD) sandhi and the object surfaces with its full citation tone. However, verbs are not the only constituents which can undergo this sandhi. In fact, beyond the verb and object phrases, Xu et al.

(1986) list the subject+verb, verb+complement, and adverb+verb boundaries all as common positions in which RD sandhi occurs, though verb+object phrases remain the most common.

While RD dominant sandhi is quite variable in its occurrence, Xu et al. also note that many words and phrases have been lexicalized as consistently using either LD or RD sandhi in any position in a sentence, while yet other phrases can alternate meaning based on which sandhi is applied; for example, application of left- versus right-dominant sandhi can lead to minimal pairs between nominal compounds and verbal phrases, as shown in 2) below, where square brackets indicate sandhi domain boundaries in the surface forms:

2) a)	/LH/	/LH/	→	[L	H]
	ts^hɔ	mi	→	ts^hɔ	mi
	'fry'	'noodle'		'fried noodles'	
b)	/LH/	/LH/	→	[44]	[LH]
	ts^hɔ	mi	→	ts^hɔ	mi
	'fry'	'noodle'		'(to) fry noodles'	

In 2a), we can see LD sandhi applying to the disyllable as a single sandhi domain, resulting in a nominal phrase, while in 2b) the same two syllables form two domains, the first of which undergoes RD sandhi, resulting in a verbal phrase.

Despite all that Xu et al.'s work has to say about RD sandhi, literature on the phenomenon since has been scarce until more recent studies. Duanmu (1995) briefly mentions RD sandhi in verb+object contexts, analyzing it as the result of additive phrasal stress to the object that causes it to be metrically more prominent than the immediately preceding verb; this imbalance in prominence leads to the tonal neutralization on the verb. However, his analysis relies on the syntactic categories of the involved syllables, and thus does nothing to explain the variable occurrence of RD sandhi in other, non-verbal contexts. Yan's (2019) dissertation on the application of LD versus RD sandhi found that beyond simple syntactic category and lexical

specification, a host of other factors can determine whether a disyllable undergoes RD sandhi. For example, lower frequency phrases and verb+object phrases with more transparent (i.e. compositional) semantics were more likely to undergo RD sandhi or no sandhi rather than LD sandhi.

Takahashi (2011, 2013) tests RD sandhi applications in real phrases at different speech rates as spoken by New Shanghainese speakers, concluding that in spite of previous impressionistic research, this pattern should be understood as phonetic reduction rather than phonological neutralization. This is due to significantly different contours being found for nearly all tones at all speech rates that maintain the same general trajectory of pitch as their non-neutralized counterparts (i.e. T1 still falls slightly, T2 still rises slightly, etc.). Zhang & Meng (2016; speakers born on average ~1990) and Ling & Liang (2018, speakers born 1983-1993) confirm these results in disyllabic verb+object contexts. Zhang & Meng (2016) however find that with nonce verb + real object pairs, speakers do not fully reduce in the way seen for real words; nevertheless, they generally did not apply RD sandhi in these contexts. While RD sandhi may not be a true sandhi in the sense of a phonological process that operates over abstract tonal categories, it nevertheless is a salient part of the Shanghainese language, especially in the somewhat more naturalistic sentences recorded in this dissertation.

2.5.5. Prosodic Studies in Shanghainese

In addition to the phonetic and phonological literature concerning the local behavior of the sandhi processes, several researchers have also investigated larger prosodic units and the ways in which focus is realized in Shanghainese. In terms of domain determination, all studies

make reference to morphological or syntactic categories and boundaries. In the case of left-dominant sandhi (LD sandhi), there seems to be a general divide between metrical (i.e. stress- or prominence-based) and purely-syntactic interpretations. In an extreme metrical approach, M. Chen (2000) radically proposes that there are no underlying tones in Shanghainese, interpreting the spreading sandhi as the result of a weight-to-prominence constraint. This constraint ensures that syllables with glottal rhymes never receive a prominent H tone in domain-initial position, as they are "light" syllables. However, this analysis explains only a small subset of the Shanghainese lexicon (i.e. compound words whose initial syllable has a glottal rhyme), ignores the falling T1 entirely, and offers no explanation of contrastive tones in monosyllables.

In another metrical analysis, Duanmu (1993, 1995, 1997) argues from data concerning long multi-morphemic compounds and mono-morphemic loanwords that the Shanghainese sandhi domain is a disyllabic trochaic foot which is aligned to the left edge of a morpheme. Sandhi domains longer than two syllables arise from incorporating degenerate feet into nearby well-formed feet with consideration of morphological boundaries such that the degenerate foot may merge with the previous or following disyllabic sandhi domain. Under this view, left-dominant (LD) sandhi is the product of two opposing constraints—one that only stressed (i.e. initial) syllables may realize their tone; another that all syllables must be realized with tone. Additionally, he shows that the morphological sensitivity of footing requires subparts of long compounds to be identical to their isolated, non-compounded forms. While Duanmu's analysis accounts for the treatment of long compound and loan words in Shanghainese, it does not concern itself with domain determination in a full sentence. It also fails to address the issue of grammatical particles that can never initiate a sandhi domain or the realization of pitch on monosyllables.

Analyses such as Zee & Maddieson (1980), Jin (1986), or Selkirk & Shen (1990) propose definitions of a sandhi domain where boundaries are determined by the syntax or morphological structure of a phrase. For Selkirk & Shen (1990), the left edges of lexical words of the categories noun, adjective, and verb project a sandhi domain edge; the domain then continues until the next defined left edge in the sentence (i.e. the next content word). This approach explains the behavior of function words and particles, as well as giving predictions for sandhi domain phrasing in full sentence utterances; however it is completely deterministic, and does not allow for variation in phrasing.

Beyond this definition of the sandhi domain, Selkirk & Shen (1990) also propose a higher level in the prosodic hierarchy, the major phrase, or MP. They argue for this grouping of sandhi domains based on two phenomena: post-focal tonal deletion and unexpected groupings of sandhi domains. The major phrase of Shanghainese, they state, is determined by the projection of a left prosodic edge from the left syntactic edge of a maximal lexical phrase XP (i.e. NP, AP, VP, etc.). The reasoning for this analysis is based on the following claim: deletion of H tones occurs within the MP that contains the sandhi domain with the focused element. We can see this effect in their schematized example, modified in 3) below. Here, italics show constituents under focus.

3) a)

[[P	[N	]$_{NP}$]$_{PP}$[V	[N	]$_{NP}$]$_{V'}$]$_{VP}$	Syntactic Structure
(	)(		)(	)	Major Phrase Boundaries
(LH	)(LH LH)	(LH	)(LH	)	Underlying tones w/ SDs
lə˨	o˨ li	ts^{h}iɪ˨	mi		

‘at home eat noodles => ‘eat noodles at home’

b)

[[P	[N	]$_{NP}$]$_{PP}$[V	[N	]$_{NP}$]$_{V'}$]$_{VP}$	Syntactic Structure
(	)(		)(	)	Major Phrase Boundaries
(LH	)(L H)	(LH	)(LH	)	Normal LD Application
lə˨	o˨ li	ts^{h}iɪ˨	mi		

‘eat noodles at home’

c)	[[P	[**N**	]$_{NP}$]$_{PP}$[V	[N	]$_{NP}$]$_{V'}$]$_{VP}$	Syntactic Structure
	(	)(		)(	)	Major Phrase Boundaries
	(LH	)(**L**	**H**)	(L	)(LH)	Focus 1
	ləʔ	***oʔ***	***li***	tsʰiəʔ	mi	

‘eat noodles at ***home***’

d)	[[P	[N	]$_{NP}$]$_{PP}$[**V**	[N	]$_{NP}$]$_{V'}$]$_{VP}$	Syntactic Structure
	(	)(		)(	)	Major Phrase Boundaries
	(LH	)(L	H)	(**LH**	)(LH)	Focus 2
	ləʔ	oʔ	li	***tsʰiəʔ***	mi	

‘***eat*** noodles at home’

In these examples, we can see the proposed MP boundaries aligned to the left edges of lexical XPs, remaining consistent through all focus conditions. Note that the prepositional phrase ‘at home’ *ləʔ oʔ li* in this analysis is considered to be an adjunct to the verb, and thus, it initiates the VP—the verb ‘eat’ *tsʰiɪʔ* does not initiate the VP itself. In 3c), we see the focus of the prepositional object *oʔ li* ‘home’ results in H tonal deletion in the following word, ‘eat,’ regardless of the intervening sandhi domain boundary observed in the non-focused context, 3b). However, in 3d), where the verb ‘eat’ is focused, the following sandhi domain ‘noodle’ is left with all tones intact, resisting post-focal tonal deletion as it initiates its own NP. Note that they do not explore other reasons for apparent H tone loss (e.g. re-phrased sandhi domains), instead simply naming it de-accenting.

Selkirk & Shen find more support for their definition of a MP in the behavior of prepositions. As stated before, non-lexical elements, such as prepositions, should form a sandhi domain with what precedes them, as can be seen in the borrowed example 4a) below. However, there are instances, such as 4b) where this generalization about sandhi domain formation is apparently violated; this unexpected parsing forms the basis of their MP-formation.

4)	a)	V	[P	[N	]$_{NP}$]$_{PP}$					Syntactic Structure
		(			)(	)				Major Phrase Boundaries
		(LH	LH		)(LH	LH)				Underlying w/ SDs
		(L	H		)(L	H)				Normal LD Application
		zz	ləʔ		zɑ̃	he				
		be	in		Shanghai					
		'to be in Shanghai'								
	b)	V	[[P	[N	]$_{NP}$]$_{PP}$	Prt	N	]$_{NP}$		Syntactic Structure
		(	)(	)(				)		Major Phrase Boundaries
		(LH	)(LH	)(LH	LH	LH)(LH	LH)			Underlying w/ SDs
		(LH	)(LH	)(L	H	L	)(L	H)		Normal LD Application
		zz	ləʔ	zɑ̃	he	ɣəʔ	ɲi	tsz		
		be	in	Shanghai		LINK	son			
		'to be the son who is in Shanghai'								

In 4a), we see the expected sandhi domain pattern: the preposition 'in' *ləʔ* forms a sandhi domain with the preceding verb, as it is not a lexical element. However, in 4b), the exact same verb and preposition form two separate sandhi domains. The main difference between these two phrases, however, lies in the syntactic structure. In 4b), the preposition 'in' *ləʔ* initiates a new NP (i.e. 'the son who is in Shanghai'), while in 4a) it begins a non-lexical PP ('in Shanghai'). Selkirk & Shen argue this unexpected prosodic structure arises from the interaction of the strict layering hypothesis and the constraints governing MP determination; the MP's need to have a boundary at the left edge of lexical XPs overrides the sandhi domain's preference. This surprising syntax-prosody mapping along with the occurrence of post-focal tonal deletion form the basis of Selkirk & Shen's MP in Shanghainese. However, Yan's (2017) dissertation, as mentioned above, finds a host of non-syntactic effects (e.g. frequency, semantic transparency) on choice of sandhi application and thus phrasing—this shows the inherently variant nature of prosodic domain determination in Shanghainese.

In more recent work, Y. Chen and colleagues have also phonetically investigated the relation between different kinds of focus, prosodic constituents, and phonetic realization of sandhi. Y. Chen (2009) thoroughly investigates the tonal correlates of focus, and challenges the claim that H tone targets are phonologically deleted post-focally. Instead, she finds that the most reliable characterization of post-focal tones is register lowering (where the entirety of the pitch range is lowered), although she also finds that pitch range compression can occur. In addition, she tested the realization of disyllables beginning with different tones under both focus[5] and contrastive topic conditions utterance-initially. The summary of her findings are presented in the following table, modified from the appendix of Y. Chen (2009):

SD-initial Tone	**Max F0**		**Min F0**		**F0 Range**	
T1 (53)	F↑	T↑	F↑		F↑	
T2 (34)	F↑	T↑		T↓		
T3 (23)	F↑	T↑	F↓		F↑	T↑
T4 (<u>55</u>)	F↑	T↑		T↑		
T5 (<u>12</u>)	F↑	T↑		T↑	F↑	T↑

Table 5. Statistically significant pitch changes of disyllables under the Focus (F) and Contrastive Topic (T) conditions (compared to Non-Contrastive Topic). ↑ indicates increased mean; ↓ indicates decreased mean. Citation tones given with tone letters copied from Table 1.

In these findings, we see complex tonal interactions between the lexical tones and the pragmatic context elicited, though there is a clear pattern of increased maximum F0 across all contexts and tones. She also notes that duration is increased in both focus and contrastive topic contexts, but this only applies to the non-glottal tones.

[5] The type of focus (e.g. broad/narrow) was not specified; in such cases, narrow focus is usually intended.

In a later study, Sun & Y. Chen (2015) explore the realization of tone and duration in disyllables broken across three different boundaries: compound-internal, verb+object, and subject+verb. With Selkirk & Shen's analysis of the MP, post-focal tonal effects should only be observed after the compound-internal boundary. Instead, Sun & Y. Chen find that it can occur in over each of the boundaries.[6] Though their general findings on f0 contours support the idea that post-focal register lowering is a phrasal marker, they do not give further characterization of their major phrase beyond saying that it "is minimally binary and consists of at least two prosodic words [i.e., sandhi domains]," and noting that focus causes the insertion of a MP boundary to the left of the focused element. Despite Sun & Y. Chen's reluctance to state the determining factors for MP-boundary establishment, Selkirk & Shen's proposal for determining MPs is effectively invalidated, as Sun & Y. Chen (2015) show tonal compression and lowering can extend over lexical XP boundaries.

In terms of focal effects on phrasing, there are mixed results. On one hand, Ling & Liang (2018) find that their five younger speakers (born 1983-1993) do not insert new sandhi domain boundaries in the middle of disyllabic LD sandhi domains, even if the second syllable is focused; nevertheless, increased pitch range and duration accompany focus. On the other hand Tian & Kuang (2020) find more variation with their seventeen speakers (no speaker age given), with three different strategies given only T1 (falling) + T2 (high rising) disyllables. In one group, when the second syllable of a disyllabic word is contrastively focused, a new sandhi domain boundary is inserted directly to its left. In another, they unexpectedly found that despite giving

[6] Note that it was not the intention of Sun & Y. Chen (2015) to confirm or refute Selkirk & Shen's (1990) proposal for the Major Phrase; rather, it seems that they expected post-focal tonal compression in each of the three boundaries from the beginning of the experimental design. However, their results still implicitly refute Selkirk & Shen's analysis.

their speakers sentential contexts that were meant to elicit focus, they did not realize any prosodically-encoded contrastive focus in the target sentence; they take this to mean that focus is not necessarily phonologically realized, even if a focused reading was intended by the speaker, though this could simply be the effect of laboratory-style speech that participants sometimes use in such experiments. In the last group of speakers, they find similarities to Ling & Liang's speakers in that there is not a clear new sandhi domain boundary to the left of a focused second syllable, but instead, the observed tonal changes are not completely attributable to normal LD sandhi application. Instead, because the second, focused syllable seems to maintain its register (high vs. low) but not its contour (rising vs. falling), they characterize this as gradient or partial sandhi application, as the tonal distinction has not been fully neutralized.

Lastly, the precursor to the current dissertation is my master's thesis (Roberts 2016), which was an initial attempt to apply an Autosegmental-Metrical framework to the tonal system of Shanghainese. The goal of the thesis was to generally understand the prosodic hierarchy, propose an inventory of boundary tones, and characterize the various phrasal effects on tone realization and sandhi application, all based on observations of pitch tracks in multiple sentence types (see Section 3.1. for a full review of methods). In terms of the prosodic hierarchy, three levels were proposed, following previous literature: the sandhi domain, the major phrase, and the intonational phrase. At the sandhi domain level, few outright claims were made, other than using a generative, two-tone inventory of L+H and H+L, staying agnostic as to the role of prominence. However, I found a theretofore unmentioned effect that I called 'cross-boundary tonal dissimilation:' if two identical tonal targets were adjacent over a sandhi domain boundary, the second, domain-initial tone would always have a more extreme realization (i.e. domain-initial highs were higher, domain-initial lows were lower)—this happened categorically with all

speakers, except with a few whose IP-initial pitch range expansion could override this, leading to a sandhi domain-final H being higher than a sandhi-domain initial H.

While the major phrase was proposed due to previous research, it was found that Selkirk & Shen's rule (left edges of lexical XPs project major phrase boundaries) was untenable—both of their diagnostics (i.e. post-focal tonal deletion and forced sandhi domain breaks from lexical XP boundaries) had ample counterexamples. Despite this, no other diagnostics for the major phrase could be identified, and so this level of phrasing was only tentatively included in the analysis. However, Selkirk & Shen's characterization of the sandhi domain behavior (left edges of lexical words project sandhi domain boundaries) generally did hold, though with the odd exception, like non-utterance-initial pronouns forming their own sandhi domain.

For the intonational phrase, I found a previously unmentioned effect of "IP-initial pitch range expansion" whereby the vast majority of IP-initial sandhi domains have the widest pitch range of any in the sentence. Though this process was not categorical for all speakers, all speakers widened their pitch ranges for at least the majority of the sentences recorded. I also proposed four IP-final boundary tones: low (L%), mid (M%), high (H%), and upstepped high (^H%). M% was the most common, unmarked IP-final boundary tone, while L% indicated some sort of negative or corrective intent on the part of the speaker. H% and ^H% were both used in questions, with ^H% specifically being used for echo questions; rather than just a simple target, it involved raising the pitch range over the entirety of the IP-final sandhi domain.

In terms of focal effects outside of the focused sandhi domain, both full post-focal tonal compression and tonal de-accenting were seen, though the reasons for using one versus the other was unclear. Additionally, in some cases, instances of *pre*-focal reduction were observed, where

tones were reduced (as part of RD sandhi), and small sandhi domains immediately pre-focus became part of larger domains to their left.

Additionally, a few other points of previous research were confirmed: my older speakers generally use a low domain-final boundary tone, though some vary and have an early inserted low target in some sentences, as described by Chen (2008); application of focus resulted in the focused sandhi domain becoming the f0 maximum of the IP (Chen 2009); and application of RD sandhi was highly variable, affecting essentially any syntactic category of word (Xu et al. 1986, Yan 2018). Additionally for RD sandhi, the degree of tonal reduction differed widely across the identified instances, with some appearing very close to their citation tones, and others where a rising tone had become falling between a previous H and a following L target, showing the truly gradient nature of this 'sandhi.' The only consistent characterization of tonal reduction was that the resulting pitch was lower than previous H tones and higher or equal with previous L tones.

Many of these processes and boundary tones will be discussed in greater detail with accompanying pitch tracks and Sh_ToBI (Shanghainese Tones and Break Indices) annotation below. In this dissertation, the inventory and behavior of the IP boundary tones is one of the biggest changes between my MA thesis and the current proposal. Additionally, I have introduced the concept of prominence to the characterization of the word-level tones themselves.

2.6. Restatement of Goals

Now that the theoretical framework and some facts about Shanghainese have been established, the goals of this dissertation will be reviewed in light of the ample background literature. In the introduction, five questions/goals were stated; they are repeated below for convenience.

i. What is the best characterization of the lexical tonal distinctions in Shanghainese?
ii. What are the prosodic units forming the phrasal hierarchy, and how can they be recognized?
iii. What is the inventory of contrastive post-lexical tones?
iv. How do tones (lexical or post-lexical) interact?
v. Create an annotation scheme for Shanghainese Tones and Break Indices (Sh-ToBI).

The first question is perhaps the hardest to give a definitive answer to. Traditional dialectologists and generative phonologists have taken different approaches to the number of tones in the language, taking either a more surface-oriented five-tone system or an abstract two-tone system, respectively. Generative, two-tone analyses depend on complementary distribution between tones and phonation, which has been called into question by the loss of phonation in younger speakers. Nevertheless, these younger speakers still consistently voice ‘breathy’ stops intervocalically and use breathiness cues in tone identification, showing they retain at least traces of the phonation contrast. At the same time, pitch seems to be the primary cue to the contrast in monosyllables for speakers of all ages. Also at debate is the domain of these tones themselves. As shown in multiple psycho-linguistic studies, Shanghainese speakers of all ages seem to store both the citation forms of individual syllables and the full ‘sandhi’ forms of disyllables, with tones already spread. Despite this dual storage, speakers are also still able to productively use the sandhi process in novel forms. These findings make it difficult to characterize one form as being more ‘surface’ or ‘underlying’ than the other, as both have representations in the lexicon of

speakers. Consequently, it seems that currently spoken Shanghainese has both two and five tones; to be both a lexical/syllable tone language and a lexical pitch accent/phrasal tone language.

The analysis developed in this dissertation itself is mixed. While it generally follows a generative or abstract approach to the individual tonal contrasts, it takes a pitch accent view to the prosodic system overall, which is more surface-oriented. These decisions are based in the autosegmental-metrical framework used to analyze the language as well as the tonal behavior encountered; however, it is important to remember the larger picture of the knowledge speakers have demonstrated. More on this conflict between 'surface' and 'underlying' representations in Shanghainese will follow in the Section 6.2.

The remaining questions are more straightforwardly assessed. In relation to the phrasal hierarchy, most researchers agree that there are three levels of phrasing above the syllable/morpheme in Shanghainese; however, there are few descriptions of how these levels of phrasing are to be recognized other than the spreading behavior of 'sandhi domains.' Roberts (2016) found that Intonational Phrases (IPs) are marked with initial pitch range expansion as well as final boundary tones, but no evidence for an intermediate level of phrasing (ip) was found—this remaining question of the ip diagnostics is answered in the model below.

In looking at the post-lexical tones in Shanghainese, there are two major places to investigate. First is the low boundary tone at the end of the Accentual Phrase (AP) and the variation between early and late low tone targets. The second is the inventory IP-final boundary tones, which Roberts (2016) argues has four contrastive tones. The characterizations of these

post-lexical tones will be discussed extensively below and developed beyond earlier claims, especially in relation to IP-final tones.

The fourth question has to do with the interaction of various tones, lexical or post-lexical. While Roberts (2016) described the difference between AP-initial and non-initial H and L targets as dissimilation, the model below recasts it as a result of a prominence imbalance. Such an analysis avoids issues with post-lexical dissimilation effects, and supports a more abstract analysis of the lexical tones in Shanghainese. While some relationships between IP-final boundary tones and lexical tones were investigated in Roberts (2016), they have largely been replaced as the inventory of IP boundary tones has been reanalyzed considerably in this dissertation.

The final point is not so much a question to be answered as a concrete product: to create a relatively efficient and useful annotation system for tonal and intonational events as well as the types of juncture in Shanghainese. This system is called Shanghainese Tones and Break Indices Labeling, or Sh_ToBI. Throughout the dissertation, the exact ways in which Sh_ToBI is similar to and differs from other ToBI models will be discussed in more depth.

3. Methodology

This chapter outlines the three datasets included in this dissertation, where speakers, stimuli, and recording situations differed. The first dataset is the same data that was used in my master's thesis (Roberts 2016, reviewed above at the end of Section 2.5.5.), while the latter two datasets are presented for the first time as a part of this dissertation. The remaining subsections will describe the differences between the three datasets before concluding with the procedure for recording and analysis that was shared by all of them.

3.1. Dataset I

The recordings from Dataset I were used in my master's thesis, and were recorded in 2015. In general, the stimuli recorded were exploratory in nature, and close work with a linguistically-naïve native speaker (speaker ID: 1a) was required in order to make a start in analyzing the language.

3.1.1. Dataset I Stimuli

The initial set of sentences was constructed in the following way: 25 words and phrases ranging from two to five syllables (note that in Chinese, each syllable is one morpheme), controlled for each initial syllable's citation tone (T1-T5) and non-initial-syllable tone (T3[7]), were assembled. Each of these 25 target words were placed first in three different basic carrier

[7] In five syllable phrases, not all non-initial syllables were T3 due to difficulty in finding such tonal shapes.

sentences that contained the word in sentence-initial, -medial, and -final position. Additionally, the words were included in sentence-final echo question positions. Together, these sentences yielded 125 highly controlled stimuli. After this, more natural sentences (as opposed to the quotation-type frame sentences, 伊讲__给侬听 'He says __ for you to hear.') were constructed using the same target words in sentence-initial, -medial, and -final position. Five-syllable phrases were left out of these sentences, resulting in another 45 stimuli. These sentences ranged from shorter, straightforward sentences (e.g. 我的鼻头疼死了 'My nose hurts.') to more complex, multi-clausal constructions (老师听说侬还无没做功课 'Teacher heard that you still haven't done your homework.'). Additionally, a set of sentences were constructed by adding an increasing number adjectives to a nominal head (e.g. 高的红的…椅子高头有张报纸 'The big, red… chair has a newspaper on it.'). Finally, a small set of six short dialogues were crafted to elicit focus of certain lexical items as well as particular pragmatic or discourse-related meanings such as correction and questioning. Both parts in these dialogues were acted out by the single speaker; the short conversations often triggered a correction of a previous word, for example:

5)	A: 余先生关照阿妈伐？ ɦy ɕìsã kuètsɔ aʔma va	'Did Mr. Yu tell grandma?'
	B: 勿是，余先生是关照**爷爷**咯。 vəʔ zz, ɦy ɕìsã zz kuètsɔ **iaia** ɣəʔ	'No, Mr. Yu told GRANDPA'

3.1.2. Dataset I Speakers

Ten speakers (2 male) recorded at various locations; six of the ten speakers' data is analyzed in this dissertation; the other 4 were not included due to either poor audio recording or seeming dialect differences from the other speakers. All speakers were recorded in person the greater Los Angeles, California region, except speaker 1a, who lives in the Kansas City metro

area and was recorded over Skype. Additionally, all speakers were surveyed for their linguistic backgrounds, and they were asked to self-evaluate their pronunciation of 12 character pairs which have undergone or resisted mergers in various Shanghainese dialects. According to Z. Chen (1999), these mergers (or lack thereof) can help define the dialect group that a speaker belongs to. No large dialectal differences were found in the speakers participated, all speakers included in this dataset(s) had merged 7-11 of the 12 pairs presented.[8]

Speaker ID	Gender	Year of Birth	Age as of 2020	Main District(s)
1a	F	1964	56	Huangpu
4a	F	1954	66	Changning
5a	F	1963	55	Changning
7a	F	1961	52	Huangpu
9a	F	1946	74	Huangpu
10a	M	1950	70	Huangpu

Table 6. List of speakers in Dataset I, with identification number, gender, year of birth, age as of 2020,[9] and the main district(s) in which the speaker lives/lived.

3.1.3. Dataset I Procedure

Speakers were recorded using Audacity v.2.1.1 (Audacity Team, 2015) on a CAD u37 cardioid condenser microphone. All stimuli were presented to speakers in Simplified Chinese characters at least twice: once in a familiarization phase, where speakers were allowed to read through all the sentence stimuli in an organized list and practice, and then again in the recording phase, where the sentence stimuli were presented one-by-one in randomized order. During the

[8] Speakers who had more mergers than others generally overlapped with speakers who had fewer; e.g. a speaker with 9 merged pairs had the same pairs merged as someone with 7 mergers, plus two. See Appendix A. for more information.

[9] Note that the speakers' ages were different at the time of recording depending on the dataset (Dataset I in 2015, Dataset II [Table 7] in 2017, and Dataset III [Table 8] in 2019). In order to ensure consistent comparison of speaker age across all years of recordings, speaker age is only reported as of 2020 in all datasets.

recording phase, speakers were encouraged to repeat the sentence as many times as necessary to achieve a fluent utterance. After recording, the best token(s) of each sentence stimulus were extracted from the full recording and displayed in Praat (Boersma & Weenink, 2015). As I was collecting the data, the process was conducted mostly in Mandarin with some English when necessary.

In general, Praat's pitch tracking tool was used to evaluate tonal patterning. In the cases where the pitch track gave errors or was absent, a narrow-band spectrogram was used to find local maximal and minimal f0 values. Additionally, in cases of tonal ambiguity, impressionistic evaluation of tonal events aided the tone labeling process, though precedence was given to actual pitch values in Hertz wherever possible. While this concludes a brief outline of the general approach used in analyzing the data, the next chapter, which outlines the proposed intonational model as a whole, more thoroughly lays out the steps taken in identifying and describing the intonational phenomena of Shanghainese.

3.2. Dataset II

In Dataset II, recorded in 2017, a much more tightly controlled list of sentences were used to confirm the cross-boundary 'dissimilatory' effects found in my master's thesis. As reviewed above, I found that when two identical tonal targets are adjacent across a sandhi domain boundary, the second, domain-initial target will always be more extreme; in other words: domain-initial highs are higher than other high targets, and domain initial lows are lower than other low targets.

3.2.1. Dataset II Stimuli

The sentence stimuli were always four domains long, and were varied by the tones present in the second and third sandhi domains, allowing for various cross-boundary tonal combinations to be tested in sentence-medial position. Stimuli were constructed for each combination of tones across the sentence-medial boundary, treating domain-final default low targets (L_{DEF}) as separate from low targets assumed at the time to be linked to lexical tones (from HL [T1] or LH [T2-3] tone target, L_{LEX}). Thus, there were six different cross-boundary tonal contexts tested (listed here with an intervening '–' indicating a sandhi domain boundary): H–H; H–L_{LEX}; L_{LEX}–H; L_{LEX}–L_{LEX}; L_{DEF}–H; L_{DEF}–L_{LEX}. While the analysis that informed the stimulus creation is not held to in this dissertation, the sentences recorded are nevertheless extremely useful in looking at cross-domain tonal behavior in Shanghainese.

3.2.2. Dataset II Speakers

Six participants were recorded (1 male), with one being excluded entirely from analysis, due to her being outside the target age range. The resulting five speakers were born between the years of 1957 and 1975. The recordings were done in person in various locations around the Los Angeles metro area. None of the speakers from Dataset I were recorded for Dataset II. Again, linguistic background and the merger survey were collected from the speakers as well; the speakers had 6-10 of the 12 mergers in the survey. Table 7 below shows a breakdown of the five speakers included from Dataset II.

Speaker ID	Gender	Year of Birth	Age as of 2020	Main District(s)
1b	F	1966	54	Xuhui
2b	F	1975	45	Luwan
4b	F	1970	50	Xuhui
5b	F	1957	63	Xuhui, Jing'an
6b	M	1966	54	Huangpu, Hongkou

Table 7. List of speakers in Dataset II, with identification number, gender, year of birth, age as of 2020, and the main district(s) in which the speaker lives/lived.

3.2.3. Dataset II Procedure

The procedure in collecting Dataset II was largely the same as in Dataset I. Speakers were allowed to familiarize themselves with the stimuli before recording them in randomized order. Speakers were encouraged to repeat the target sentences until fluent during recording. Instead of a single repetition, however, each sentence was recorded three times. All data was analyzed using Praat, as before. Again, the data was collected in a mixed Mandarin/English setting by me.

3.3. Dataset III

Dataset III was recorded in Summer 2019 while in Shanghai, China, either at the phonetics lab at Fudan University or in speakers' residences, always in person.

3.3.1. Dataset III Stimuli

The stimuli here varied by both syntactic structure and phonological length, with less attention paid to tonal specifications. Sentences were varied by the overall structure, including standard subject-verb-object, fronted object constructions, if/then conditionals, serial and modal verb constructions, and relative clauses. Beyond general sentence structure, syntactic categories within the sentences were varied as well—determiners, numerals, measure words, possessors, adverbs all appeared in addition to the necessary nouns, verbs, and function words.

Of course, with syntactic variation in the stimuli comes variation in length; more complex constituents will necessarily be longer. However, phonological length was still varied by making each syntactic constituent (i.e. noun, number, verb, possessor, etc.) longer or shorter. With all these variables taken into account, the stimuli totaled 180 sentences, ranging from trisyllabic SVO sentences (我骂伊 'I'm scolding him'), to thirteen-syllable if/then statements (假使侬有晨光，应该跟伊讲闲话 'If you have time, [you] should chat with him.').

3.3.2. Dataset III Speakers

A total of 17 speakers (7 males) were recorded in this experiment. One speaker was excluded entirely from analysis due to difficulty reading the sentences with fluency. Due to time constraints, the analysis presented in this dissertation accounts for 10 (4 males) of the 16 speakers recorded. These 10 speakers were specifically chosen by looking at those speakers with the highest quality recordings, fewest disfluencies, and most dialectal mergers to ensure similarity. Table 8 shows speaker information for the 10 speakers included from Dataset III.

Speaker ID	Gender	Year of Birth	Age as of 2020	Main District(s)
1c	M	1970	50	Hongkou
2c	M	1963	57	Jing'an
3c	F	1970	50	Yangpu
6c	F	1970	50	Hongkou
7c	F	1974	46	Hongkou
8c	M	1937	83	Pudong
9c	F	1974	46	Huangpu
13c	M	1968	52	Huangpu
15c	F	1962	58	Huangpu, Changning

Table 8. List of included speakers from Dataset III, with identification number, gender, year of birth, age as of 2020, and the main district(s) in which the speaker lives/lived.

3.3.3. Dataset III Procedure

While the previous datasets were collected via working with the speakers in a mix of Mandarin and English, Dataset III was collected by a young, linguistically-aware native speaker of Shanghainese, who was able to conduct the entire process in Shanghainese while I was present. In addition to the regular familiarization/practice phase of stimulus presentation, the speakers went through two production phases: the first at a ‘normal’ speech rate, and the second at ‘fast’ speech rate, both determined by the speaker themselves. As there was wide variation in what counted as ‘normal’ and ‘fast’ speech rates between speakers, the results are presented in an aggregated manner in this dissertation.

Again, Praat was used for visual analysis of tonal events for comparison with behavior in earlier datasets. However, Dataset III was also used to develop the break system (ToBI) of the model presented here. This was done in close consultation with my linguistically-trained native speaker assistant in order to access intuitions about pauses and disfluencies. By comparing the

amount of perceived juncture between every adjacent syllables in a fluent sentence, a basic system of break labels was developed. After diagnostics were determined for each of these break labels, less fluent sentences were investigated to further develop the system into one that can handle many types of disfluencies, pauses, and mismatches between expected tonal behavior and perceived juncture.

3.4. Summary of Methodology

In total, 33 speakers (10 males) were recorded, with data from 21 speakers (6 males) analyzed and accounted for in this dissertation. Included speakers were born between 1937 and 1975, with an average age of 57 years as of 2020. While most data was collected from speakers living in the US using a combination of Mandarin and English in the recording session, the largest single dataset was recorded in Shanghai using Shanghainese consistently during the process. Praat and native speaker intuitions were used extensively in the analysis of the prosodic effects present in the data. Complete lists of all speakers recorded and all stimuli are included in Appendices A and B respectively.

4. An Autosegmental Metrical Model of Shanghainese Tone and Intonation

4.1. Introduction to Revised AM Model and Sh_ToBI

The shift to treating the language as a pitch-accent language comes with many changes to the tonal inventory. Instead of five or two tones, I am proposing that Shanghainese has three contrastive pitch accents: L*+H, H*, and L*, whose properties will be further described below. Important to note here is that all the pitch accents have a 'starred' tone—this diacritic indicates that the tone will always align to the prominent syllable in a given phrase. In Shanghainese, this is predictably the phrase-initial syllable, whose pitch accent always overrides subsequent syllable's pitch accents (at least in phrases that have been generated by the speaker rather than simply accessed). The cross-boundary effects uncovered in my master's thesis, originally described as dissimilatory, have also been re-interpreted as another realization of the prominence imbalance between syllables in a phrase.

Additionally, these three pitch accents are accompanied by one of two AP-final boundary tones, one rising (LHa) and one low (La/L:a). As stated before, there seem to be two realizations of the low boundary tone. One is as an AP-final low preceded by interpolation (the early low, or 'older' pattern), labeled as La. The other involves a low plateau from the first non-tone-bearing syllable in the AP until the end of the phrase (the late low, or 'younger' pattern), labeled below as L:a. While the difference in realization is not contrastive, it is not currently predictable where one form will occur versus the other, and because of the interest in the behavior of this boundary tone in the literature, I have included a way to annotate the difference so that future researchers have a way to identify, find, and analyze it. When the pitch accent is a singular tone (H*, L*), it

must be accompanied by an AP boundary tone at all times (La/L:a in the case of H*, LHa in the case of L*), even in monosyllables. For the one bitonal pitch accent, L*+H, the AP boundary tone La/L:a can only occur in phrases longer than three syllables.

As in my master's thesis, there are three levels of phrasing, though in order to bring the terminology describing Shanghainese more in line with traditional AM models, some phrases have be renamed. For instance, the sandhi domain, the lowest phrasal level, will be referred to as an Accentual Phrase (AP), as it is the domain of the pitch accent. The major phrase of Selkirk & Shen (1990) will be referred to as an intermediate phrase (ip), again as a way to harmonize the Shanghainese and intonational literature. The last phrasal level, the intonational phrase (IP) luckily needs no such renaming. The following figure schematizes the intonational system of Shanghainese, showing the hierarchy of prosodic units and the tonal inventory defining each prosodic unit as proposed in this dissertation.

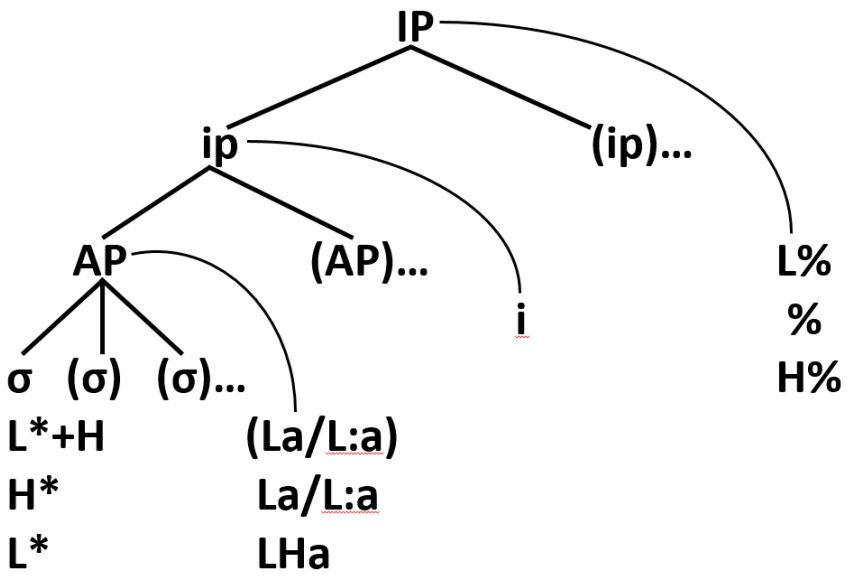

Figure 1. The intonational system of Shanghainese, showing phrasal hierarchy, pitch accents, and boundary tones. Straight lines indicate constituency relationships; curved lines indicate boundary marking phenomena. IP = intonational phrase; ip = intermediate phrase; AP = accentual phrase.

Here we can see the proposal for the intonational hierarchy of Shanghainese: the Accentual Phrase (AP) must contain exactly one of the three pitch accents (L*+H, H*, L*), with its corresponding AP-final boundary tone (La/L:a or LHa). As in other autosegmental-metrical models, a 'starred' tone, X*, indicates tones that are aligned to prominent syllables, while the 'plus', +X, indicates tones that will align to the syllable following the prominent syllable if the AP is longer than one syllable. Lastly, Xa indicates a boundary tone that align to the end of an AP. Here, the parentheses around (La/L:a) specifically indicate that the Low boundary tone only appears in APs longer than two syllables when following a L*+H accent. The intermediate phrase (ip) in turn must contain at least one AP and is marked with a final 'i;' while the ip does not have a final boundary tone, it is nevertheless distinguishable via segmental lengthening and pitch reset, which the 'i' represents. Lastly the Intonational Phrase (IP) must contain at least one ip, and must occur with one of the three IP-final boundary tones, L%, toneless %, and H%, which are also accompanied by extensive segmental lengthening and pitch range compression.

Shanghainese Tones and Break Indices Labeling (Sh_ToBI) was developed for this dissertation in order to annotate and analyze tonal contours. All sound files are presented with Sh_ToBI labeling in a visual format, with both spectrograms and pitch track visible. Under this, Sh_ToBI has seven tiers: Tones, IPA, Hanzi, English, Breaks, Transl., and Misc.

The Tones tier is a point tier that contains all the labels for pitch accents, boundary tones/events, as well as focus and tonal reduction. For the pitch accents, the tone labels are aligned to the appropriate f0 (pitch) maxima or minima, so long as they occur on their host syllable. However, there are also cases of delayed tonal target realization, where an f0 minimum

or maximum occurs later than expected but is nevertheless perceptually linked to its regular host syllable; in such cases, a '<' annotation is used under the surface tonal target while the tone marker is aligned to the phonological host syllable, indicating that the phonetic tonal event actually belongs to the immediately preceding phonological target. In cases where a tonal delay is not perceptually linked to its phonological host, no '<' would be used, and a disfluency would be marked. Tones can also be pushed earlier in an AP when other IP-final tones crowd the space—in these cases '>' will be used to show a surface tonal target that is realized earlier than its phonological host. For all boundary tones (i.e. La/L:a, LHa, i, L%, %, H%), the annotation is always aligned to the end of the phrase-final syllable. The two kinds of focus will be described more in depth below, but their annotations also appear on the Tones tier as either 'aF' (narrow focus) or 'aC' (corrective focus)—these marks are aligned to the beginning of focused accentual phrases. Lastly, tonal reduction of any kind is shown via parentheses '( )' around the annotation for the affected tone.

The remaining tiers are more self-explanatory. The IPA tier is an interval tier that gives a broad surface transcription of each syllable/morpheme in the sound file in IPA. All segmental transcriptions are followed immediately (i.e. with no space) by the citation tone number for the morpheme, 1-5. The Hanzi tier is an interval tier that gives a syllable-by-syllable transliteration of the Shanghainese utterance into Simplified Chinese characters.[10] The English tier gives a word-by-word (*not* by syllable) gloss of the sentence—compound words are usually glossed whole, while grammatical particles are labeled separately from their surrounding morphemes in all capital letters. Under the English tier is the Breaks tier, which includes the break indices, a

[10] Note that phonetic stand-in or historically inaccurate yet semantically correct characters may be used to improve readability for Chinese speakers of any dialect. See Appendix B for full list of stimuli as they were presented to speakers.

numerical indication of perceived juncture between every syllable (see Chapter 5 for the proposed break indices). Next, the Transl. tier (short for translation) gives an idiomatic English translation, labelled with a single point aligned to the center of the sentence being translated. Lastly, the Misc. tier is a point tier used for miscellaneous labels indicating background noise, laughter, filler word/segments, or other annotator comments; in some cases where the Tones tier becomes too crowded for easy reading, some tone annotations may be moved to the Misc. tier with a note about moving for readability. The Misc. tier is not included in all labeled pitch tracks given below, as in most utterances selected for presentation in this dissertation have no labels appropriate for the Misc. tier.

4.2. Pitch Accents and the Accentual Phrase

This section will briefly outline the basic behavior of the pitch accents within APs, while cross-AP-boundary interactions of the tones will be discussed in more detail in the following sections. The table below summarizes the multiple analyses of contrastive pitch in Shanghainese, allowing for easy comparison between my newly proposed system and earlier accounts:

Impressionistic Monosyllabic / Citation Realization (Xu et al. 1986)	Voiced Onset?	Glottal Rhyme?	Tradi-tional No.	Selkirk & Shen (1990)	Standard Generative (e.g. Duanmu 1997)	Roberts (2016)	**Roberts (current)**
[53]	N	N	**T1**	/HL/	/HL/	H+L	**H* La/L:a**
[34]	N	N	**T2**	/MH/	/LH/	L+H	**L*+H (La/L:a)**
[23]	**Y**	N	**T3**	/LM/			
[44]	N	**Y**	**T4**	/MH/			
[12]	**Y**	**Y**	**T5**	/LM/			**L* LHa**
[33]~[31]~[11]	N/A	N/A	Default Low	/L/	/L/	L-	N/A

Table 9. Table comparing different analyses of Shanghainese citation tones as well as the default low that is inserted in domains longer than two syllables.

The H* pitch accent (always paired with La/L:a) corresponds to the falling T1 in the traditional system, or /HL/ in a standard generative account. In monosyllables, the full falling contour is realized, while in longer APs, the La/L:a aligns to the end of the phrase. In the example shown in Fig. 2, there are four disyllabic APs, all with H* accents, all followed by some low boundary tone. This figure displays pitch tracks of Shanghainese utterances annotated in Sh_ToBI. The top part of the figure displays a pitch track on top of spectrogram of the utterance, aligned with labels on six tiers shown below the pitch track. The first (top) tier, called 'Tone' shows pitch accents and boundary tones of a prosodic unit (AP, ip, and IP). The second tier, called 'IPA', gives a broad surface transliteration of each syllable into IPA, followed by the syllable's traditional tone number (1-5). The third tier, called 'Hanzi', gives the transliteration of the Shanghainese syllables into Simplified Chinese characters. The fourth tier, called 'English', gives a more literal, word-by-word gloss of the Shanghainese sentence. The fifth tier, called

‘Break’, shows the break indices perceived by the labeler between each pair of syllables. Lastly, the final tier, called ‘Transl’, gives an idiomatic translation of the Shanghainese sentence into English. The same format is used for other pitch track figures in this dissertation (see Section 4.1. immediately above for a more detailed description of Sh_ToBI tiers).

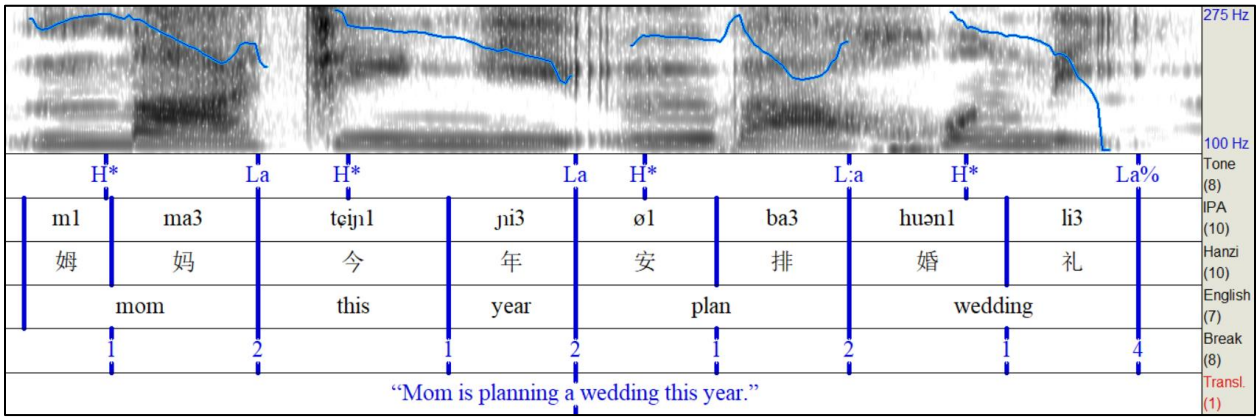

Figure 2. Annotated Shanghainese sentence ‘Mom is planning a wedding this year,’ with only H* pitch accents, but showing both La and L:a. Speaker: 4b.

In the example in Fig. 2, it appears that the speaker has used the late-target La in almost all of the APs—this is generally the case in disyllabic H*-initial APs. However, in this case it looks as though the third AP, 安排 ‘plan’, has a low target that is realized earlier in the second syllable than in any other AP, so it has been annotated with L:a instead. Thus, it appears that at least for some speakers, L:a is an option following H* even in disyllabic APs. However, the difference between late-target La and early-target L:a is more evident in longer APs, like those in Figs. 3 and 4 below.

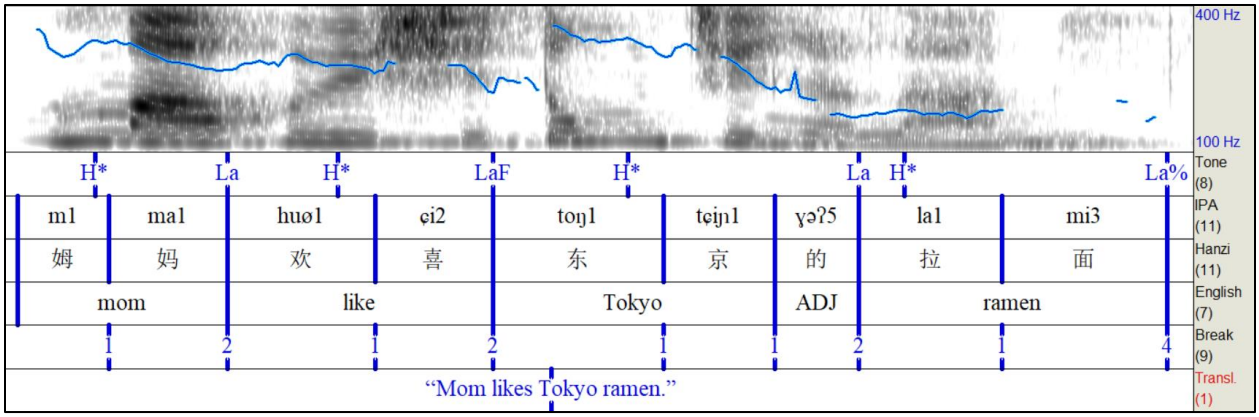

Figure 3. Annotated Shanghainese sentence 'Mom likes Tokyo ramen,' showing a late-target La at the end of the focused, third AP, 东京的 'Tokyo ADJ.' Note that 拉面'ramen' has a compressed and lowered pitch range, as the preceding AP is focused. Speaker 5b.

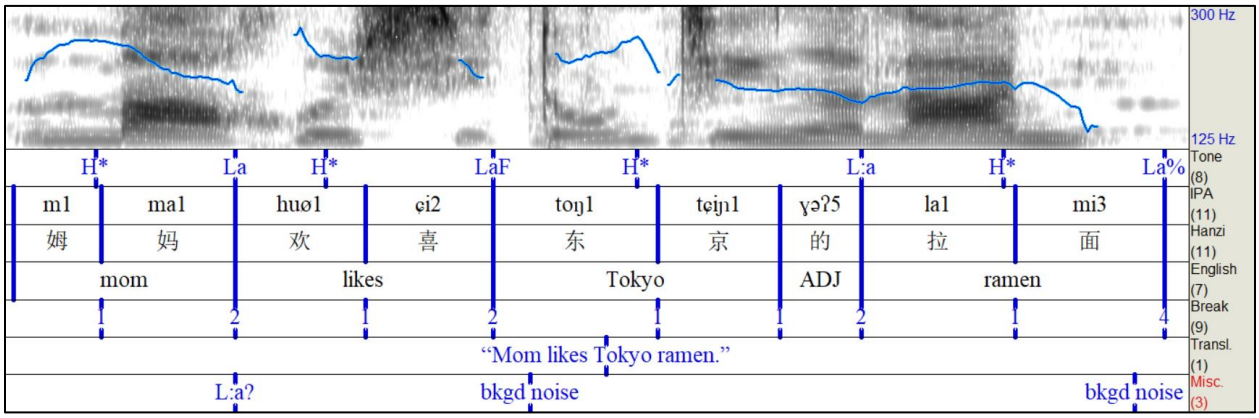

Figure 4. Annotated Shanghainese sentence 'Mom likes Tokyo ramen,' showing early-target L:a preceded by a low plateau at the end of the focused AP, 东京的 'Tokyo ADJ.' Speaker: 4b.

In both cases, a low boundary tone is present at the end of the phrase, but in Fig XX, the tone spreads leftwards until the first syllable without an assigned tone in the AP, resulting in a low plateau. Note that while in both sentences the target AP is focused, there is no relationship between focus and early/late AP boundary lows.

The L*+H pitch accent corresponds to the traditional tones T2, T3, T4, and in some cases, T5; in a generative account, this tone would correspond to the almost all of the instances of the rising /LH/ tone. While the pitch accent itself is generally rising, it can be affected by laryngeal setting on the initial syllable (T2 and T4 are high/voiceless register, while T3 and T5 are low/voiced). The contour is generally fully realized in monosyllables, with the exception of T4, where the monosyllable is high and level due to a voiceless onset and a glottal coda. In disyllabic or longer domains, the two tone targets will be split over the first two syllables. The tonal displacement pattern of T5 is not represented by L*+H, specifically due to the difference in behavior of that pitch event. In Figs. 3 and 4 below, we can compare the high (T2) and low (T3) variants of L*+H. Both pitch tracks are given in with the same pitch range view (100Hz-300Hz) and are uttered by the same speaker for ease of comparison.

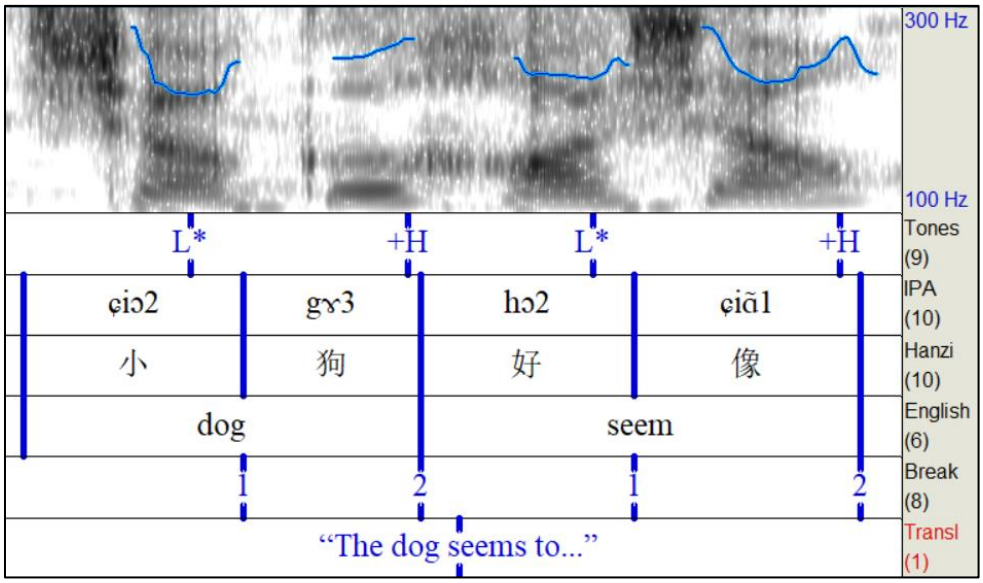

Figure 5. Annotated Shanghainese phrase 'The dog seems to…,' showcasing the high variant of the L*+H pitch accent (i.e. traditional T2). Speaker: 4b.

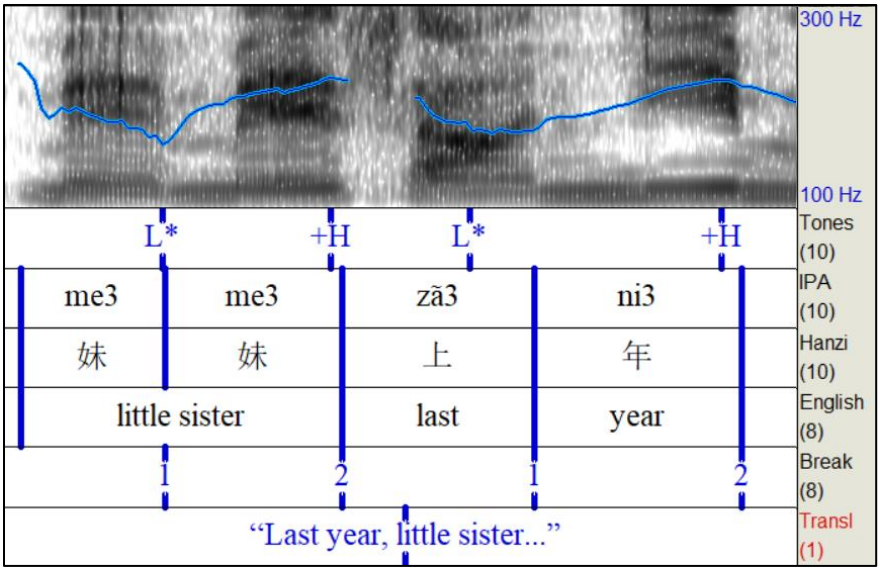

Figure 6. Annotated Shanghainese phrase 'Last year, little sister…,' showcasing the low variant of the L*+H pitch accent (i.e. traditional T3). Speaker: 4b.

Both of these pitch accents occur with La/L:a, the low AP-final boundary tone; however it only appears in L*+H-initial APs when the phrase is longer than two syllables. Fig. 7 below shows a late-target La aligned to the very end of the tri-syllabic AP 爱人的 'spouse's'. Here, the La is flanked by both a +H from its own AP, and the H* of the following H*+L monosyllable, 猫 'cat,' clearly showing the La target.

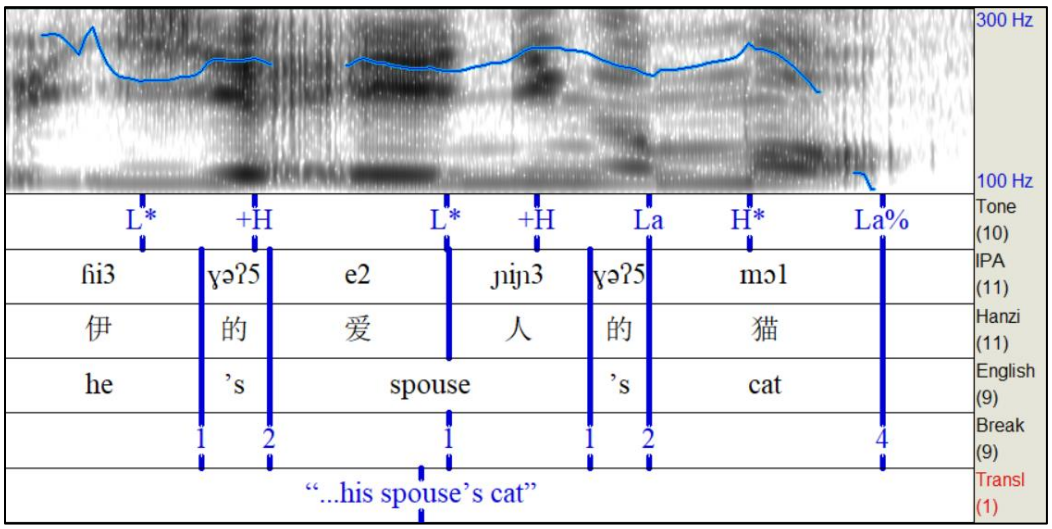

Figure 7. Annotated Shanghainese phrase '…his spouse's cat' showing AP-final La boundary tone following +H and preceding H*. Speaker: 7c.

Again, the variation between late-target La and early-target L:a is best seen in longer APs, like the word ‘fluffy’ 毛茸茸的 in the following Figs. 8 and 9, where the same sentence was uttered differently by two speakers.

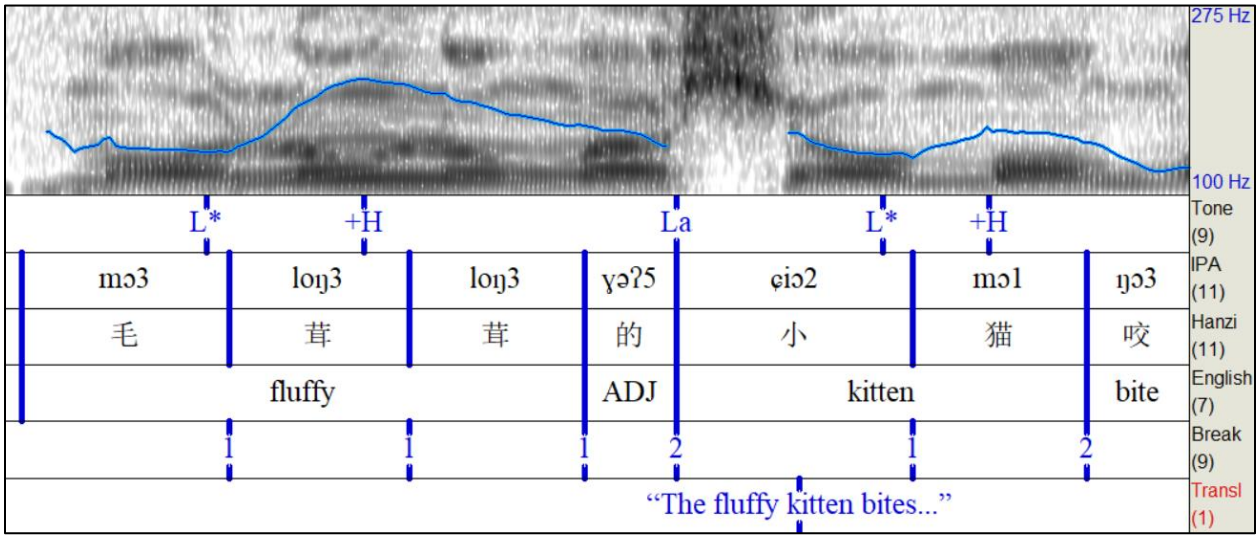

Figure 8. Annotated Shanghainese phrase ‘The fluffy kitten bites…’ showing a late-target La at the end of the initial AP 毛茸茸的 ‘fluffy.’ Speaker: 2c.

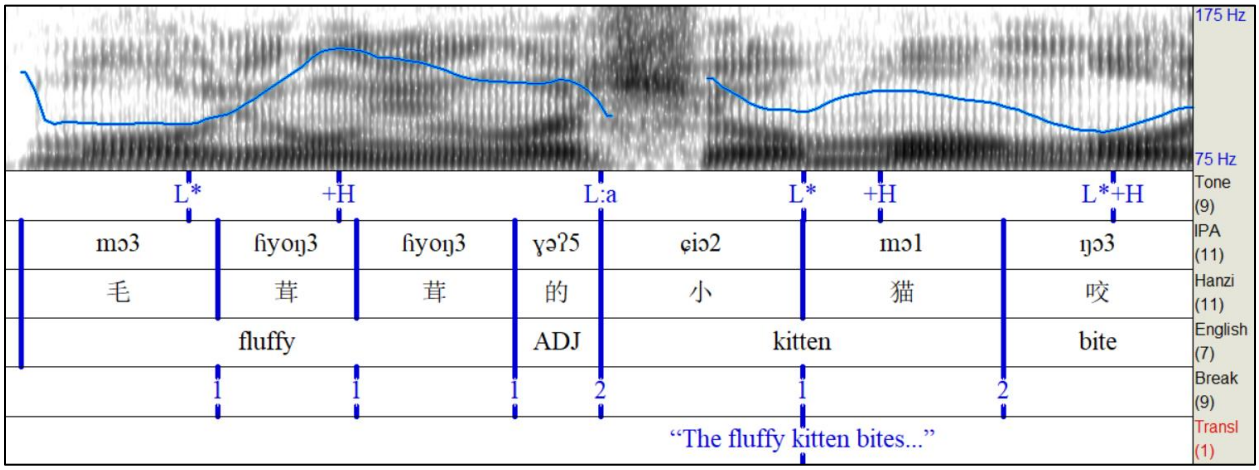

Figure 9. Annotated Shanghainese phrase ‘The fluffy kitten bites…’ showing an early-target L:a on the last two syllables of the initial AP, 毛茸茸的 ‘fluffy.’ Speaker: 1c.

Just as with H*, the La here is preceded by interpolation from the previous +H target in Fig. 8, while in Fig. 9, the L:a spreads leftwards from the end of the phrase until the first tonally unspecified syllable, resulting in a plateau.

The last pitch accent is L*, always paired with LHa, which corresponds to T5 words which show the unique 'tonal extension' pattern. Just as with H* and La/L:a, both the L* pitch accent and its LHa boundary tone is always realized, even in monosyllables. In longer APs, LHa aligns to the end of the phrase, preceded by a low plateau from the previous L*.

This pattern has been reported as potentially variable by previous research, with little agreement on the causes. According to a free variation view, while an AP-initial syllable may be listed in a traditional dictionary as T5 (by virtue of its segmental content), its AP may have either L*+H (La/L:a) or L* LHa realization in practice. However, in the data analyzed for this dissertation, a different pattern emerged. Some speakers varied in their use of L*+H or L* LHa for across different T5-initial APs, but no speaker showed variation within a single lexical item or phrase—in other words, my speakers appeared to have a phonemic split, with certain words stored as either L*+H (merging with T2-T4) or L* LHa (maintaining the unique T5 pattern), though this is yet to be conclusively confirmed.[11] In the examples below, we can see two speakers that consistently produced the same word with different pitch accents.

[11] Note: as of yet, there are no minimal pairs for any individual speaker where the segments are the same, but the words differ only in L*+H vs. L* LHa.

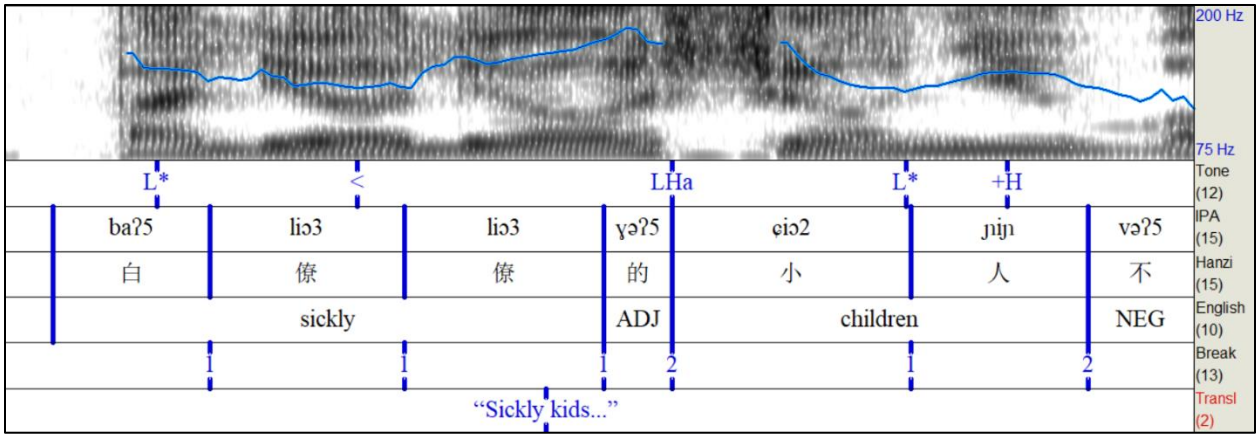

Figure 10. Annotated Shanghainese phrase 'sickly kids don't…' showing L* LHa on the T5-initial, quadrisyllabic AP. Speaker: 9a

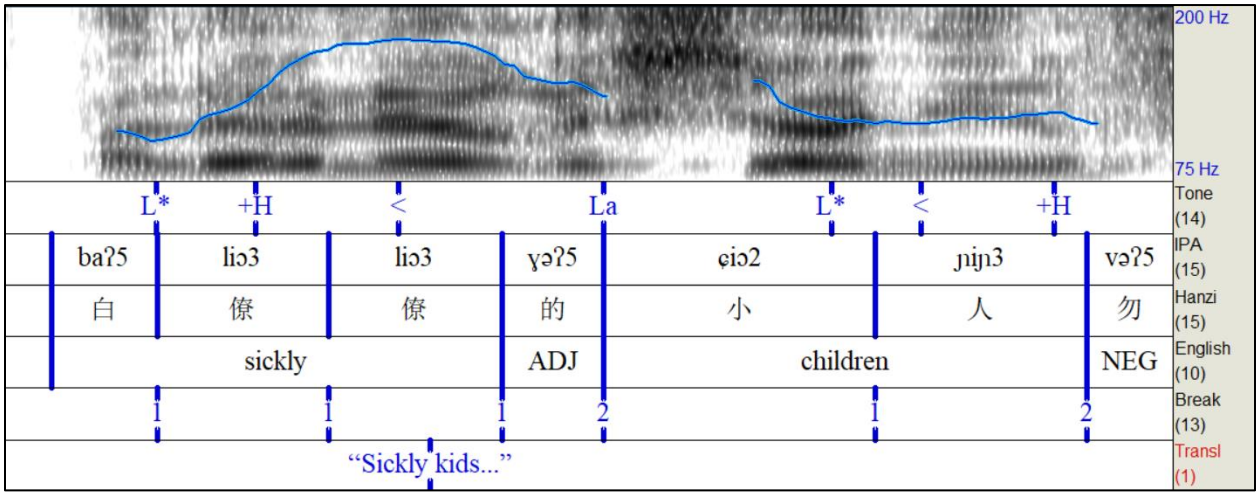

Figure 11. Annotated Shanghainese phrase 'sickly kids don't…' showing L*+H La on the T5-initial, quadrisyllabic AP. Speaker: 10a

In the Fig. 10, the L* stays on the initial syllable while a rising contour is realized on the domain-final syllable (here, it begins a bit before, as the final syllable in this AP is a short grammatical particle 的 'ADJ'). On the other hand, in Fig. 11, the speaker used L*+H La for the same lexical item, matching the tonal pattern seen in Fig. 7.

4.3. Cross-AP Boundary Tonal Behavior

In Roberts (2016), I introduced the difference in behavior between AP-initial tones and non-AP-initial tones. The general pattern is that AP-initial tones are almost always more extreme realizations of their low or high tone than their non-initial counter parts. I originally characterized this as phonetic tonal dissimilation. Such an analysis would go against the general understanding of the Obligatory Contour Principle as being part of lexical rather than post-lexical phonology.

Here, I propose a different understanding of these facts, where the concept of prominence, or phonological privilege, takes center stage. Under the analysis proposed here, H* and L* are both prominent tones; this ensures their tonal targets are always the local maximum or minimum, respectively. On the other hand, the remaining non-AP-initial tones (including both +H and the AP-final boundary tones) do not have the phonological strength to have consistent realizations, and thus their apparent targets change dependent on context. This can occur whether the tones involved are in monosyllabic or multisyllabic APs. In this section, I show the effect of prominence on L* and H* tones, and lay out the contexts where +H and La[12] can lose their discernable tone targets before identical prominent tones (before H* and L*, respectively). Only when speakers produce the early low L:a pattern, a low AP-boundary tone is visible before L*.

[12] Due to a lack of data, the non-prominent boundary tone LHa is not included here, though I predict that its H target should function just as +H does and would be subsumed by a following H*.

Looking first at the high tones, H* and +H, we can see the prominence imbalance realized in the following figure, where a disyllabic L*+H-initial AP is followed by an H*-initial AP.

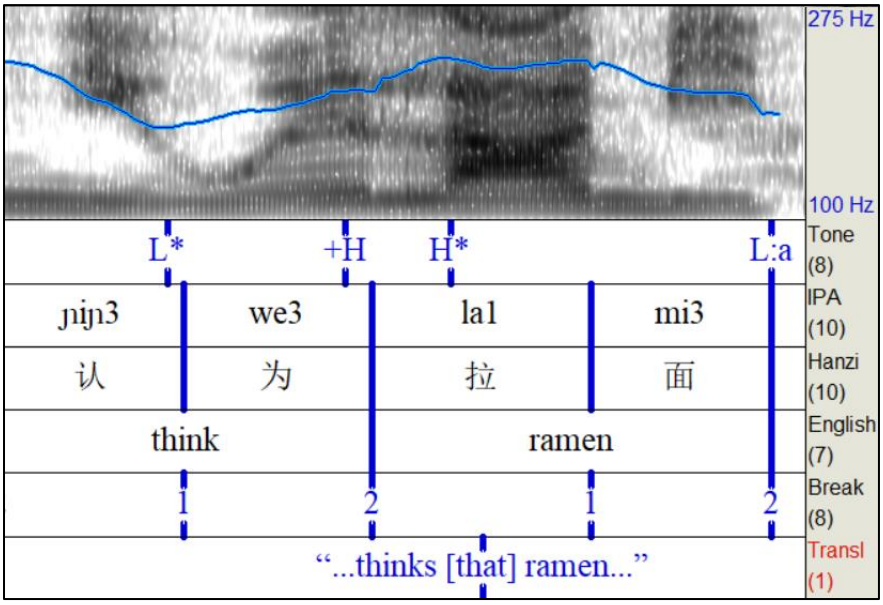

Figure 12. Annotated Shanghainese phrase '…thinks [that] ramen…,' showing the prominent H* subsuming the weak +H target between 认为 'think' and 拉面 'ramen.' Speaker: 4b.

The pitch track over the two middle APs 认为 'think' and 拉面 'ramen' is the focus here. Where a separate +H tone should appear on 为 *we*, instead there is tonal interpolation from the L* on 认 *ɲiŋ* to the H* on 拉 *la*, with no clear pitch target for +H. The prominence of H* ensures that its tone is realized as the local maximum, subsuming the previous +H target entirely.

As for the low tones (L*, La/L:a), a similar pattern can be seen when adjacent across an AP boundary. L* is always realized as the local pitch minimum, regardless of both the laryngeal setting on the initial consonant and the pitch accent preceding the La/L:a in question. In the following examples (Figs. 13-15), all La/L:a targets are preceded by L*+H. In Fig. 13, we can

see a trisyllabic AP 马路浪 ‘on the street’ starting with L*+H and ending with La preceding the next, low-variant L*-initial AP 买了 ‘bought.’

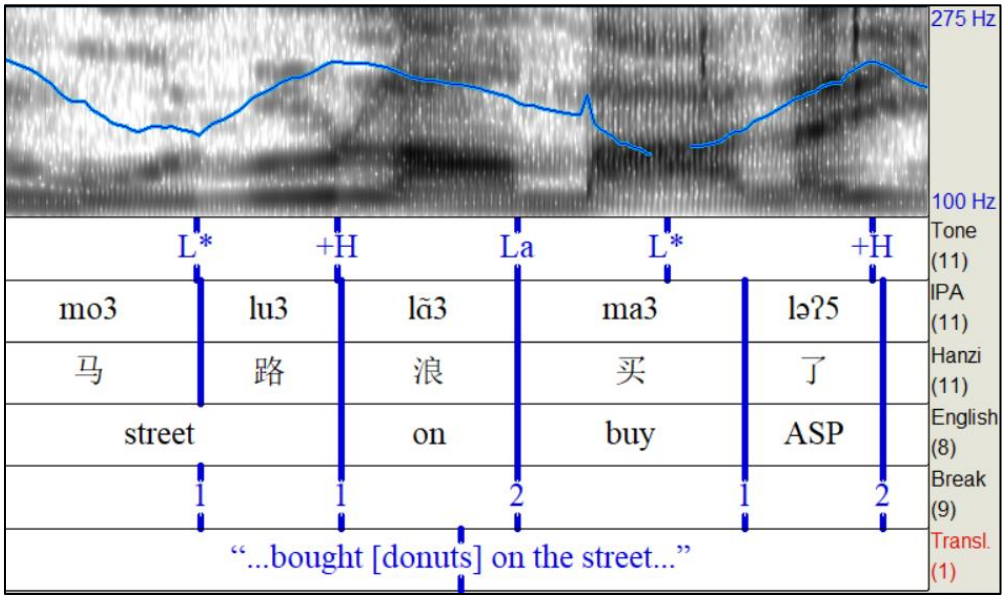

Figure 13. Annotated Shanghainese phrase ‘…bought [donuts] on the street…’ showing prominent low-variant L* subsuming the weak La target between 浪 ‘on’ and 买 ‘buy.’ Speaker: 4b.

In this example, the La target that is predicted to appear at the end of the phrase 马路浪 ‘on the street’ is completely subsumed by the following L*, appearing as normal interpolation from +H to L*. While it may seem that there is no need for a La label here, recall Fig. 7 where the same pitch accent has a La target when followed by a H*; thus, in the name of internal consistency, a La is labelled here as well, regardless of its lack of identifiable tonal target.

The pitch track in Fig. 13 shows a predicted La preceding a low-variant L*+H (i.e. T3), which triggers the complete loss of a tonal target for La. However, the same happens even when a high variant L*+H (non-glottal T2 and glottal T4). In Fig. 14, we see an La subsumed by the following high-variant L* target.

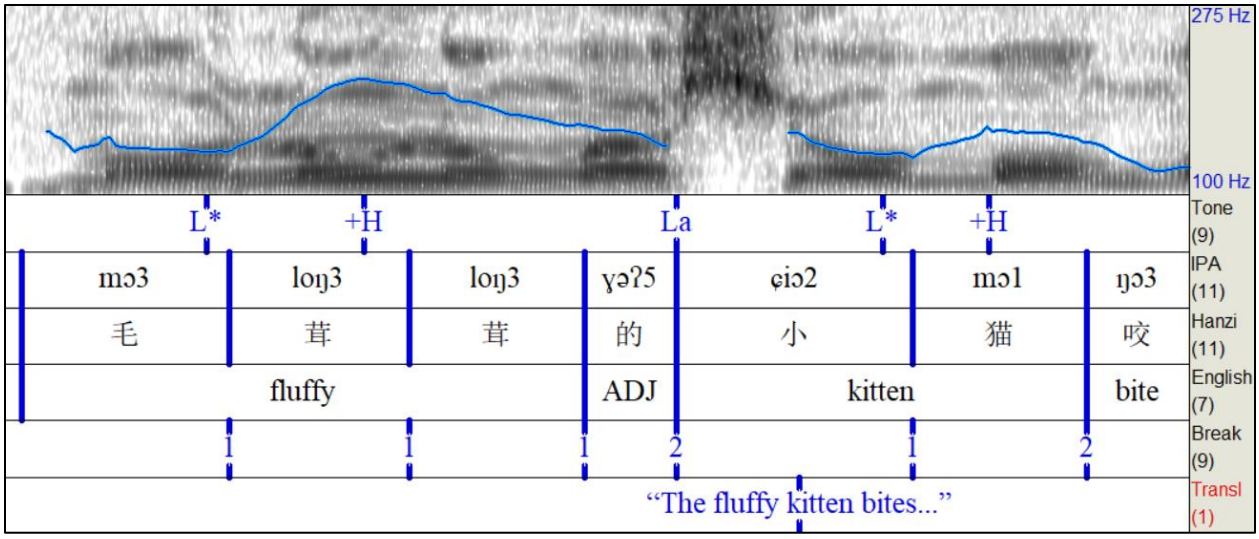

Figure 14. Annotated Shanghainese phrase 'The fluffy kitten bites…' showing prominent high-variant L* subsuming a weak La target between 的 'ADJ' and 小猫 'kitten.' Same as Fig. 8 above. Speaker: 2c.

In this sentence (repeated from Fig. 8 above), we can tell there is a late target La from the interpolation from the previous +H; however, when compared to the following high-variant L*, we see the La target again subsumed by the following L*. On the other hand, when an early-target L:a is used, then the low target becomes apparent via the accompanying plateau. Regardless, L* is still the local minimum, as evidence in Fig. 15 below.

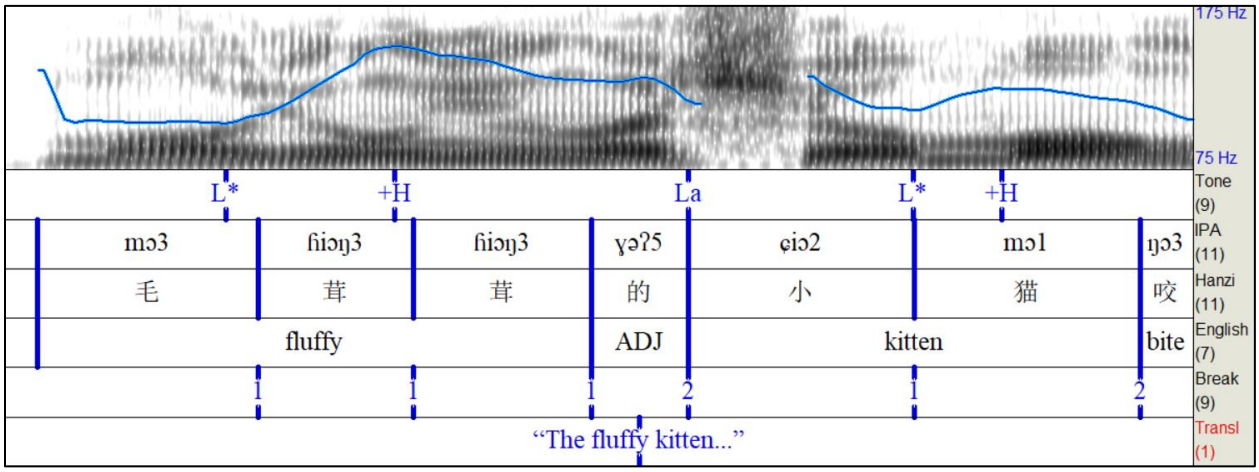

Figure 15. Annotated Shanghainese phrase 'The fluffy kitten…' showing an early low La target beginning on the second 茸 [ɦyoŋ] before a high-variant L* on 小猫 'kitten.' Same as Fig. 9 above. Speaker: 1c.

The pitch track in Fig. 15 appears similar to the younger speaker pattern of early L tone realization noted by Chen (2008) and Takahashi (2013). While in this case, the speaker in Fig. 14 (born 1963) is older than the speaker in Fig. 15 (born 1970), there is no discernable correlation between age and the behavior of La in my data overall. However, this could potentially still be explained as a result of sound change, with different older speakers having different amounts of interaction with younger speakers and different levels of personal malleability, leading to muddled age effects.

The next few figures (Figs. 16-21) show that the prominence imbalance between L* and La/L:a has the same behavior even when preceded by H* rather than L*+H. In the Fig. 16 and 17 below, we see La losing its distinct pitch target before L*, again regardless of the laryngeal setting of the prominent syllable's onset.

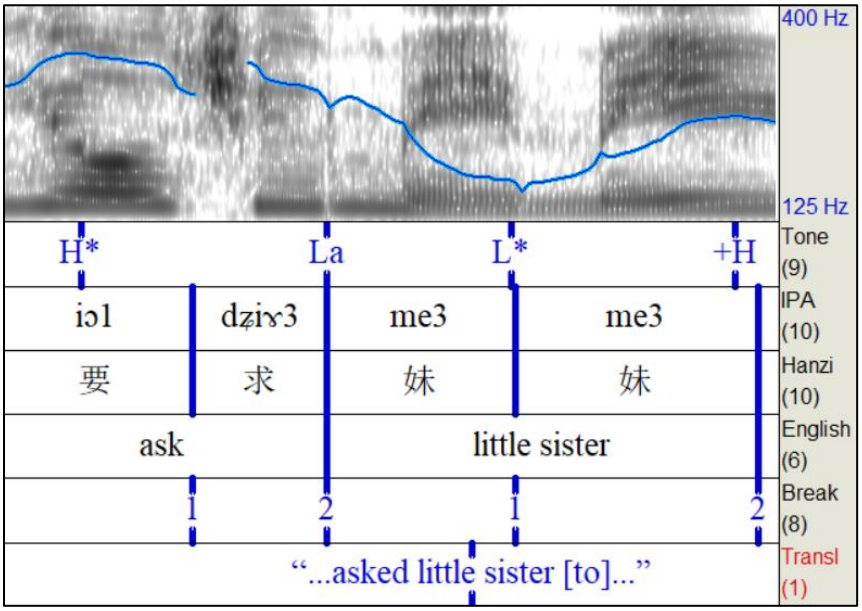

Figure 16. Annotated Shanghainese phrase '…asked little sister…' showing prominent, low-variant L* on 妹妹 'little sister' subsuming the weak +L target on 要求 'ask.' Speaker: 1b.

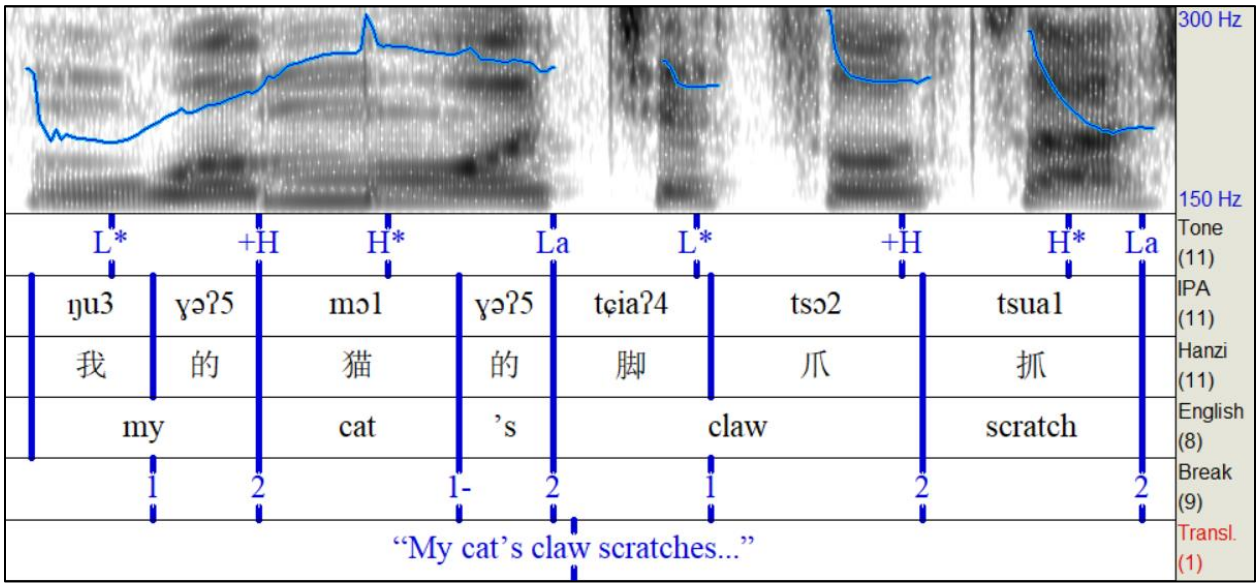

Figure 17. Annotated Shanghainese phrase ‘My cat’s claw scratches…’ showing prominent, high-variant L* on 脚爪 ‘claw’ subsuming the weak La target on 猫的 ‘cat’s.’ Speaker: 7c.

In Fig. 16, we see an H*-initial AP paired with La that precedes a low-variant L*, while in Fig. 17 we have the high-variant L* instead. In neither case is there a clear tonal target for La. It is important to note that the high-variant L* in Fig. 17 is the highest variant possible (traditional T4, voiceless and glottal); despite this, the La target is still missing, showing just how much non-prominent low tones will adjust to ensure the local pitch minimum is realized on an AP-initial L*.

Again, even in longer H*-initial APs, we can see the same interaction between prominent and non-prominent tones with both La and L:a. In the following figures, we see two speakers who have said the same sentence, with different low AP-boundary tones before a high-variant L*.

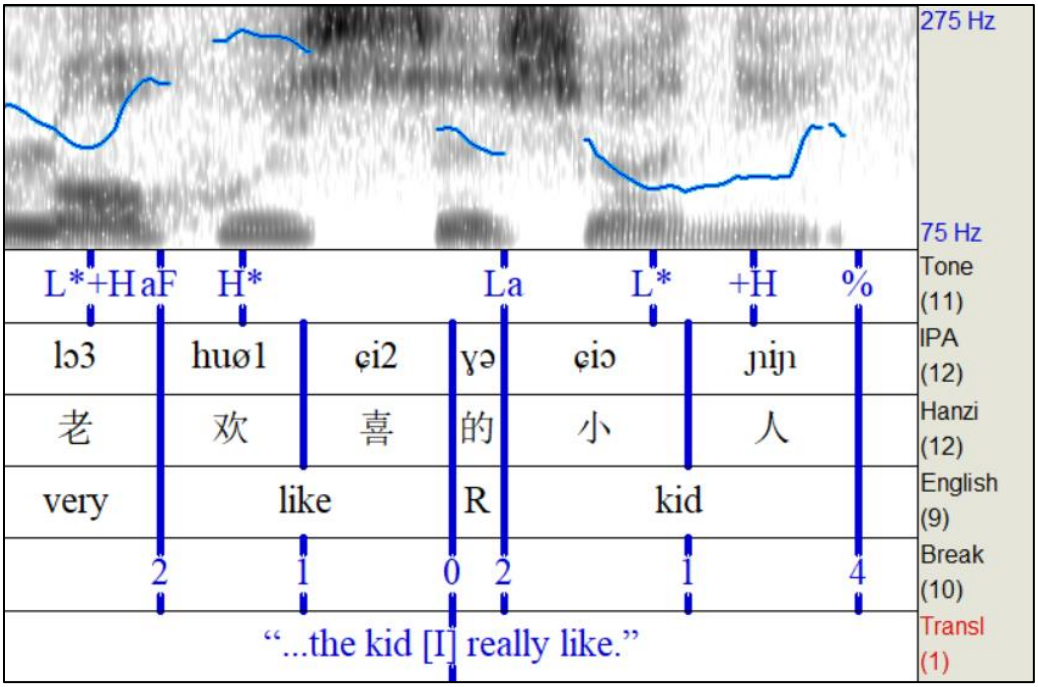

Figure 18. Annotated Shanghainese phrase '…the kid [I really] like,' showing interpolation from the H* on 欢喜 'like' to the high-variant L* on 小人 'kid,' subsuming the intervening La target. Speaker: 6c.

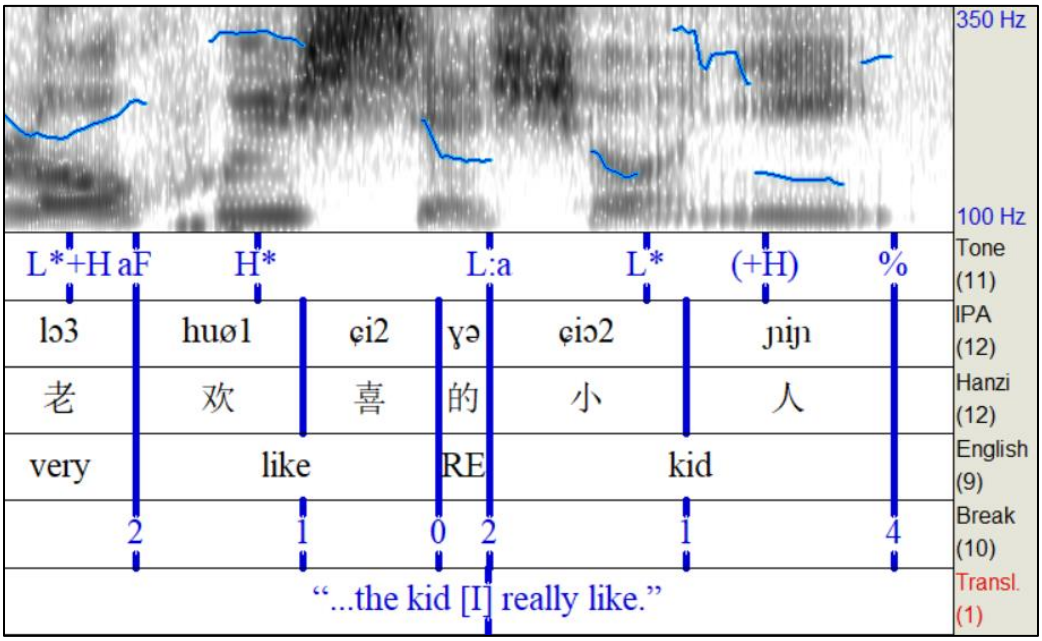

Figure 19. Annotated Shanghainese phrase '…the kid [I really] like,' showing an early-target L:a 欢喜 'like,' resulting in a short plateau before falling to the high-variant L* on 小人 'kid,' Speaker: 7c

In Fig. 18, we see the disappearance of the La target, and instead there is interpolation from the H* on 欢喜 ‘like’ to the high-variant L* on 小人 ‘kid.’ In Fig. 19, a different speaker has an earlier low target, resulting in a low plateau over the last two syllables of the first visible AP, 欢喜的 ‘like REL.’ Figs. 20 and 21 below show the same effects before a low-variant L* on the word 男人 ‘man.’

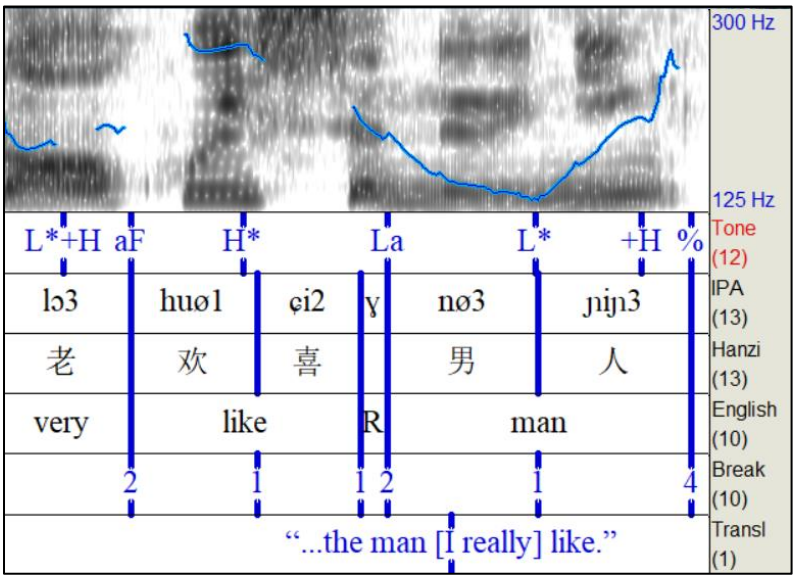

Figure 20. Annotated Shanghainese phrase ‘…the man [I really] like’ showing interpolation from the H* on 欢喜 ‘like’ to the low-variant L* on 男人 ‘man,’ subsuming the intervening La target. Speaker: 3c.

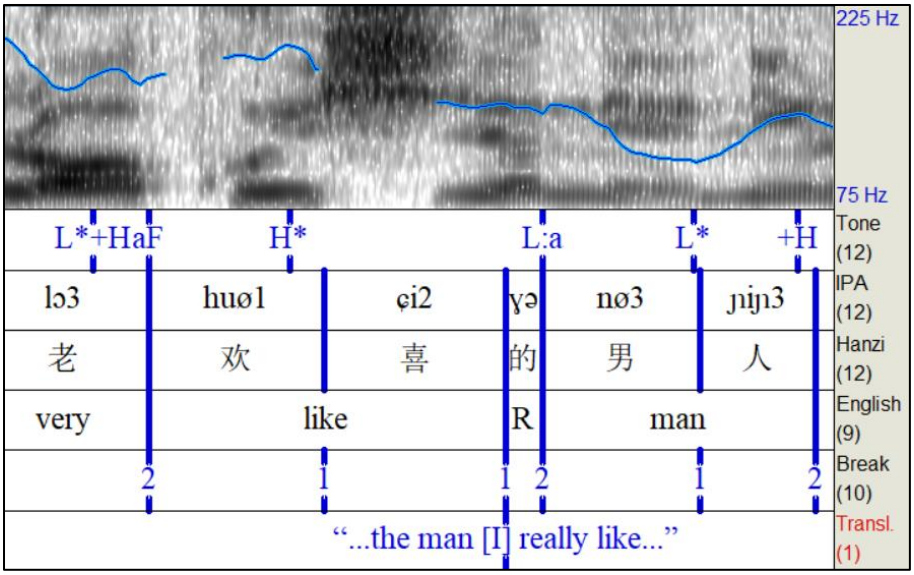

Figure 21. Annotated Shanghainese phrase 'the man [I really] like… showing an early low target for +L on 欢喜 'like,' resulting in a short plateau before falling to the low-variant L* on 男人 'man.' Speaker: 2c.

Thus we can see, regardless of high or low identity, prominent targets will always be the local pitch maximum or minimum. Such tonal behavior ensures there is never a cross-boundary pitch plateau; this may have the effect of lending salience to the left edge of AP for listeners. As seen earlier, there is variation in the realization of AP boundary lows; speakers even vary within themselves as to the production of La versus L:a; for instance, compare Fig. 14 (with interpolation) with Fig. 21 (with early low)—both of these sentences were uttered by the same speaker, 2c, in the same normal speech rate block.

4.4. Tonal Reduction

While identified as right-dominant sandhi in the majority of traditional and phonological studies on Shanghainese (Xu et al 1986, Duanmu 1993, M. Chen 2000), more recent phonetic studies suggest that rather than a categorical replacement of tone, the phenomenon is best

characterized as phonetic tonal reduction (Takahashi 2011, Zhang & Meng 2016). Despite this gradient application of reduction, the process nevertheless retains aspects of a phonological process, as it seems sensitive to syntactic structure, and specific words and phrases are stored as always undergoing reduction (Xu et al 1986, Zhang & Meng 2016). Yan (2017) confirms these and further shows that application of tonal reduction is dependent on frequency of collocation and semantic transparency of compounds and phrases. Overall, perhaps the clearest tendency in variation is that monosyllabic verbs are the most common targets of reduction.

In the data examined for the dissertation, there is no variation in lexical reduction specification, word frequency, or semantic transparency; rather the main variables manipulated were syntactic structure, tonal content, and/or phonological length. Thus, no stimuli were specifically created to elicit reduction. Despite the lack of explicit reduction contexts, all speakers produced tonal reduction, though some speakers did have lower rates of reduction in their speech, opting instead for full tonal realization. In the following section, the data show that tonal reduction may occur in multiple syntactic contexts and that there is both inter- and intra-speaker variation in its application. In terms of annotation, tonal reduction is shown in this system by putting the reduced tone in parentheses; e.g. (L*+H).

In the first example (Figs. 22 and 23 below), we can compare the apparent tones on the monosyllabic verb 来 ‘come’ in the portion of the sentence ‘… coming to Shanghai tomorrow.’

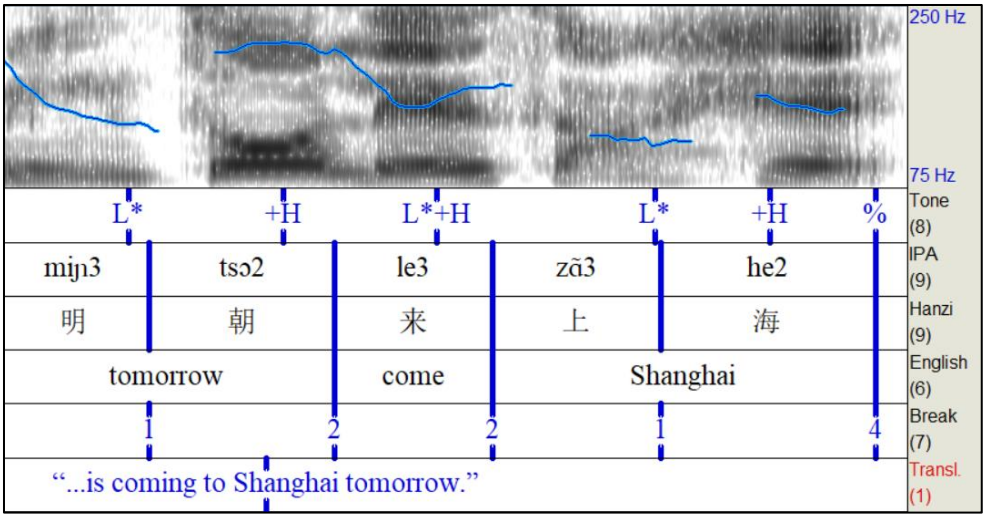

Figure 22. Annotated Shanghainese phrase '…coming to Shanghai tomorrow' showing expected L*+H contour on the monosyllabic verb 来 'come.' Speaker: 3c.

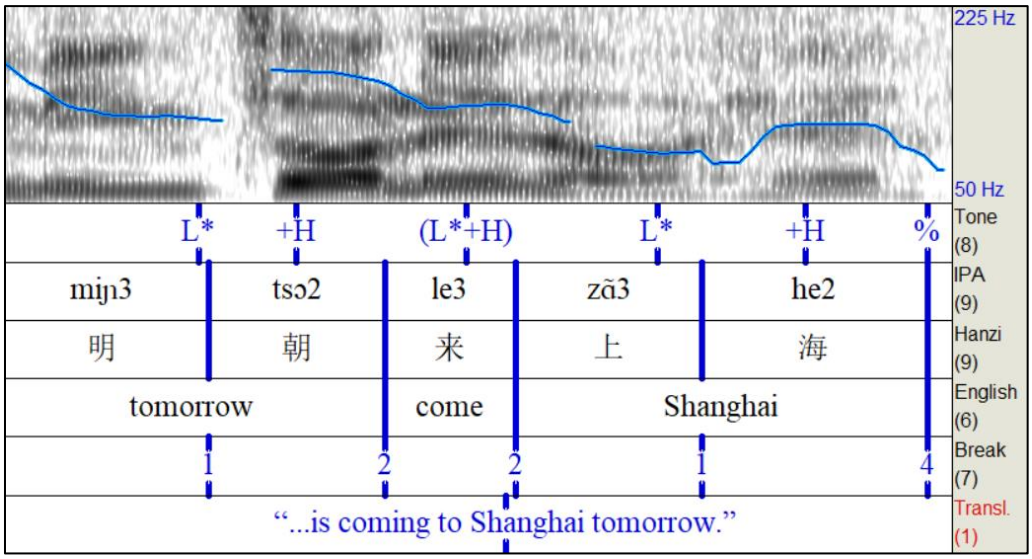

Figure 23. Annotated Shanghainese phrase '…coming to Shanghai tomorrow' showing reduced L*+H target on the monosyllabic verb 来 'come.' Speaker: 2c.

In Fig. 22, we see a normal monosyllabic L*+H tone on the verb 来 'come,' with both tone targets clearly realized; however, in Fig. 23, we see that the same verb has a nearly level pitch track, with only a hint of rise over the course of the syllable. Comparing it to both the previous and following APs, which are also low-variant L*+H, we can see that the verb 来

‘come’ has both a ‘raised’ L* and a ‘lowered’ +H, showing that reduction applies to a syllable, not just a single tone target. We can also tell that this syllable does not belong to the previous or following AP by the tonal behavior around it. In the case that the verb is the last syllable in a trisyllabic AP 明朝来 ‘tomorrow come,’ we could expect a La ‘target’ aligned to the end of the verb, which should be realized as interpolation from the previous +H on 明朝 ‘tomorrow’ to the L* on 上海 ‘Shanghai.’ No such interpolation is visible. While one could argue that the speaker produced L:a on the last syllable of the purported AP, 来 ‘come’, the perception of my native speaker consultant was that the juncture between 明朝 ‘morning’ and 来 ‘come’ was too large to be AP-internal (thus it is marked with the Break 2 index underneath). On the other hand, because the first syllable of 上海 ‘Shanghai’ is not +H, we know that the verb 来 ‘come’ does not initiate its own multi-syllabic AP. Thus we can only conclude that this is an example of tonal reduction, and in this case, it is not unexpected. While ‘come to Shanghai’ is not necessarily a lexicalized phrase, it does contain a monosyllabic verb, which makes it a normal target for this process.

However, in the next two figures, we will see tonal reduction apply to a non-traditional target. While in Xu & Tang (1988) mentions that any part of speech may undergo reduction, in subsequent research nouns have not been investigated as potential. However, in my data, it was not uncommon to find speakers who reduced nominal monosyllables, even in non-lexicalized instances.

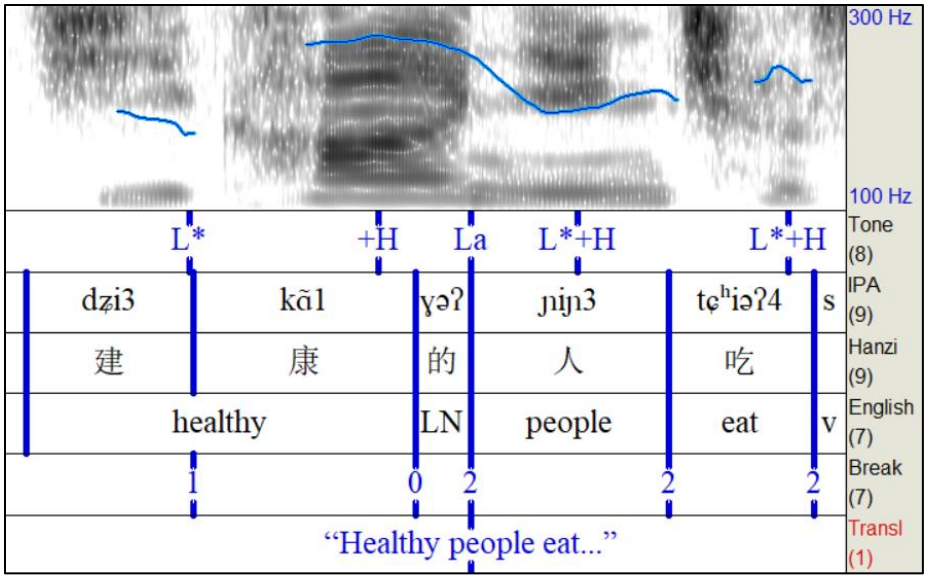

Figure 24. Annotated Shanghainese phrase 'Healthy people eat…' showing expected L*+H contour on monosyllabic noun 人 'people.' Speaker: 7c.[13]

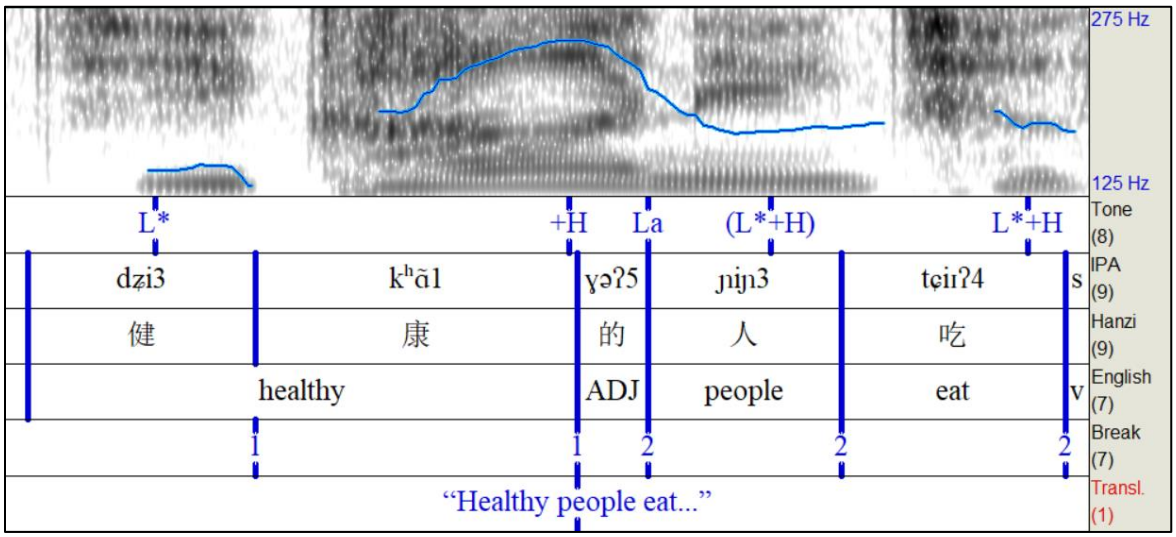

Figure 25. Annotated Shanghainese phrase 'Healthy people eat…' showing reduced L*+H contour on monosyllabic noun 人 'people.' Speaker: 6c.

In Fig. 24 above, we see a normal L*+H contour on the monosyllable 人 'person,' with both tonal targets fully realized. In Fig. 25, however, we see the same reduction as in Fig. 23 on

[13] The expected interpolation from +H to L* does not seem to occur here—this has been attributed to the 0 break, which will be discussed in Section 5.1.

the noun. While not necessarily surprising from a phonological standpoint, as it looks just like other reduced tones, the fact that this is a normal content noun not in a compound suggests that tonal reduction may have a wider usage in running speech than previously assumed.

Xu & Tang (1988) describe reduction applying to final syllables in multi-syllabic APs, but to my knowledge, there has been no investigation of this phenomenon since. While some instances may be attributable to pre- or post-focal effects (as described in Section 4.5. on focus), there are also some instances of tonal reduction that appeared in disyllabic APs, as in Fig. 26 below.

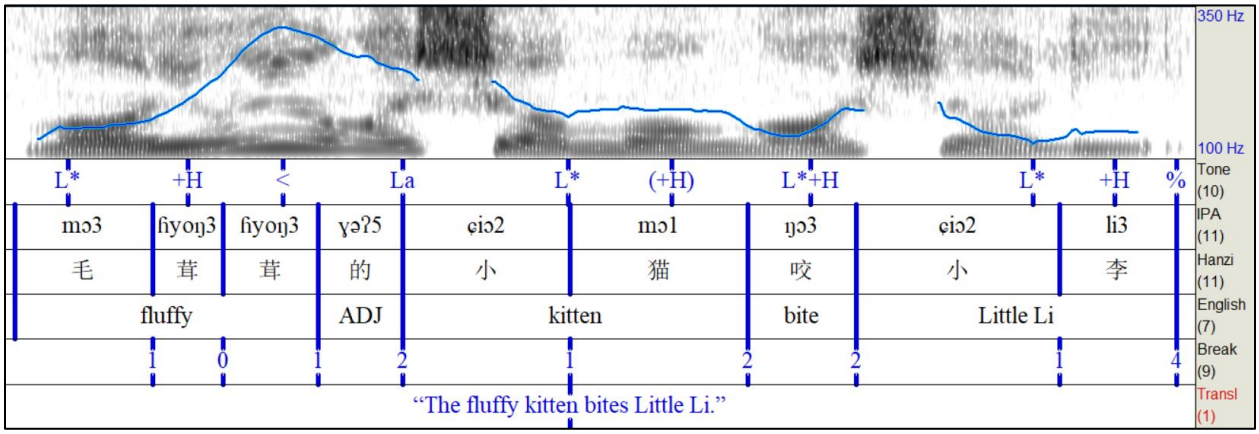

Figure 26. Annotated Shanghainese sentence "The fluffy kitten bites Little Li," showing reduction of +H on the second syllable of the AP 小猫 'kitten.' Speaker: 6c.

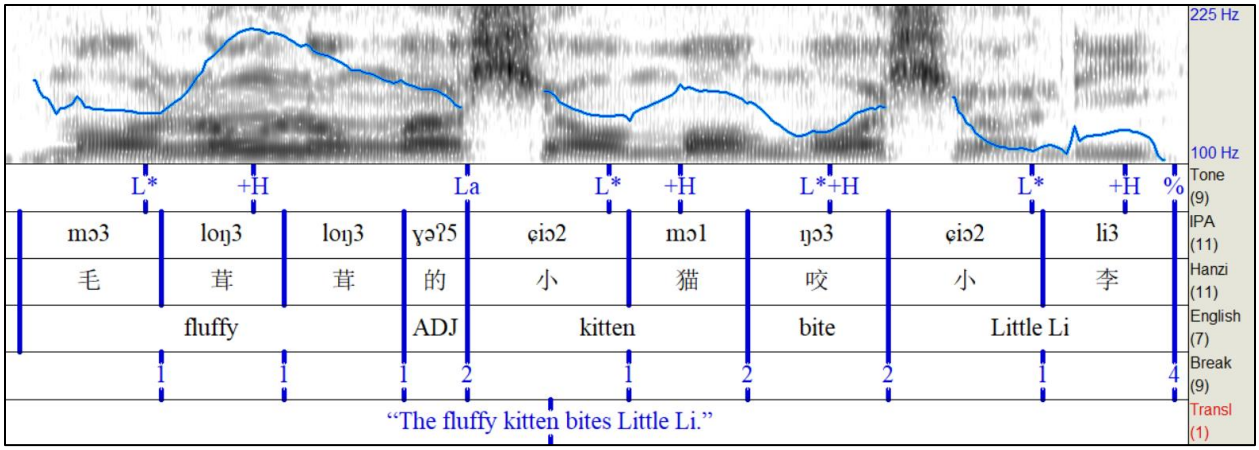

Figure 27. Annotated Shanghainese sentence "The fluffy kitten bites Little Li," showing normal production of +H on the second syllable of 小猫 'kitten.' Same utterance as Fig. 8. Speaker: 2c.

In Fig. 26, the second syllable of 小猫 'kitten' is predicted to have a +H tone, as is evident in Fig. 27. Instead, it is quite level, and only marginally higher than the preceding L*. While tonal reduction may occur in focal contexts and near phrase boundaries, as described below, neither of these seem to be present in Fig. 26. Again, comparing Figs. 26 and 27, the only major difference between them is the presence of a clear +H on 小猫 'kitten.' Thus, it seems that the relative flattening of this syllable is due to an application of the tonal reduction in multisyllabic domains as described by Xu & Tang (1988). Again, patterns of tonal reduction seem much more common in my data than in previous studies and warrant further investigation.

4.5. Identifying Focus and Testing the Intermediate Phrase

Recall that in Selkirk & Shen's 1990 (S&S) work on Shanghainese prosody, they assume a middle level of phrasing between the sandhi domain (the accentual phrase, or AP, in this analysis) and the intonational phrase (the IP), which they major phrase, and I call the intermediate phrase (ip). S&S gives an algorithm that predicts XP boundaries given a syntactic

structure: the left edge of a lexical XP projects the left edge of a major phrase, continuing up until the next lexical XP edge. They also propose that the major phrase is the domain of de-accenting, or tonal deletion, whereby all APs within the major phrase that contains the focused element lose their pitch targets on post-focal words. According to this analysis, a major phrase that follows a de-accented one should resume having regular pitch targets, resisting de-accenting. Beyond de-accenting, they provide no other phonetic indicators of their major phrase.

Thus, in order to test their claims about the existence of some intermediate level of phrasing, we must look to focus data, as S&S provide no other instance where the major phrase would have a discernable effect on the intonation or prosody. In this section, S&S's claims about the intermediate phrase will be shown to be false, from the perspective of both phrasing determination and post-focal de-accenting. Then clear instances of ip boundaries will be investigated, with a list of diagnostics for identification given.

Focus, beyond being perceptually salient, can be reliably identified by its increase in maximum f0 over the focused AP, often making it the highest pitch point in the sentence. In some cases, low targets can also be lower, increasing overall pitch range. Contra Selkirk & Shen (1990; S&S), we will see that post-focal de-accenting (i.e. tone target deletion) is not general to all kinds of focus, but rather that de-accenting only follows *corrective focus*, where a speaker is correcting a previous mistake from another speaker. Despite the existence of post-focal de-accenting in these contexts, S&S's predictions are still not born out in the data here. In more

general *narrow focus*, no de-accenting is observed at all; instead, post-focal APs have heavily compressed and lowered pitch ranges, though the degree of this compression is variable.[14]

Despite the fact that narrow focus stimuli were not explicitly designed on my part, speakers nevertheless produced sentences with emphasis in particular places, perhaps due to their own ability or willingness to 'act' in the recordings. The first few examples here are from such contexts, where speakers themselves have decided to add focus. In the first pair of examples here, Figs. 28 and 29, we see the same sentence with narrow focus on the word 根本 'simply' produced by two different speakers. The differences in tonal behavior are striking considering the predictions of S&S. In the system proposed here, narrow focus is annotated with an aF (showing the boundary of the AP, **a**, followed by a **F**ocus marker) that is aligned to the beginning of the focused AP.

[14] This aligns with the distinction between 'phonologically' and 'phonetically' de-accented strings made by Jun (2011). In phonological de-accenting, tones are deleted at the phonological level, and thus their targets never appear, while in phonetically de-accented strings, there is only post-lexical pitch range compression. While it is possible that S&S treated both forms as the same singular phenomenon of 'de-accenting,' their predictions are still not valid for my data.

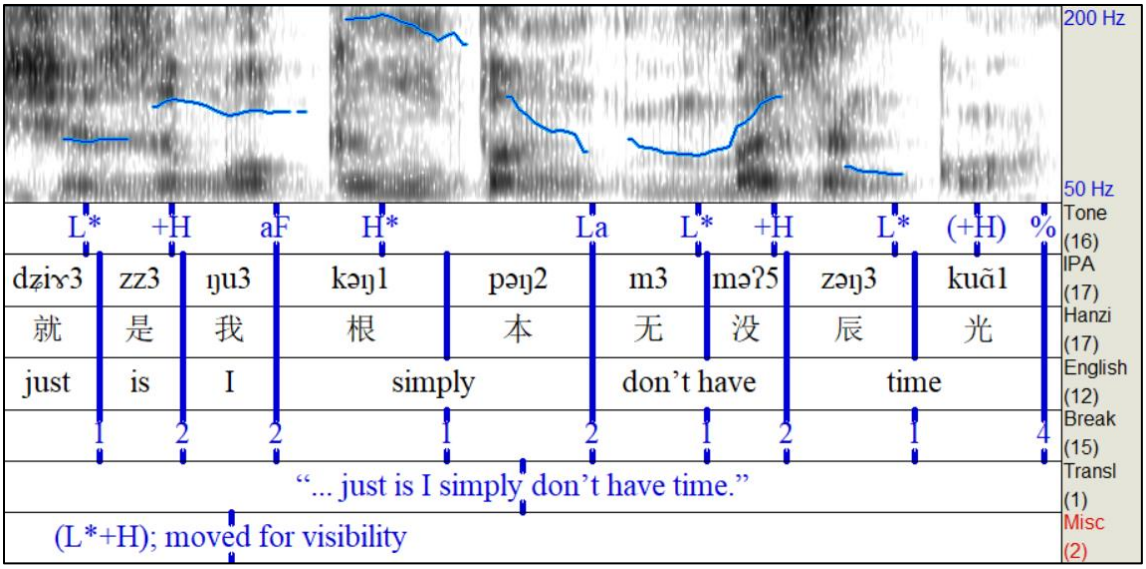

Figure 28. Annotated Shanghainese phrase '…just is I simply don't have time,' showing effects of narrow focus on 根本 'basically;' in this case, a raised H* target, reduction of preceding AP 我 'I,' and no post-focal de-accenting. Speaker: 10a.

Taking first Fig. 28, we can see the obviously raised H* target on the first syllable of the focused element, 根本 'simply,' which is the highest in the sentence. We can also see that the preceding monosyllabic AP, 我 'I' is reduced to a fairly level tone (note that the reduced tone marker has been moved to the Misc. tier for visibility in Fig. 28). Because of the very flat tone, and preceding larger juncture, it can be assumed that 我 'I' forms a reduced monosyllabic AP on its own. This can be contrasted with the same sentence said by another speaker below:

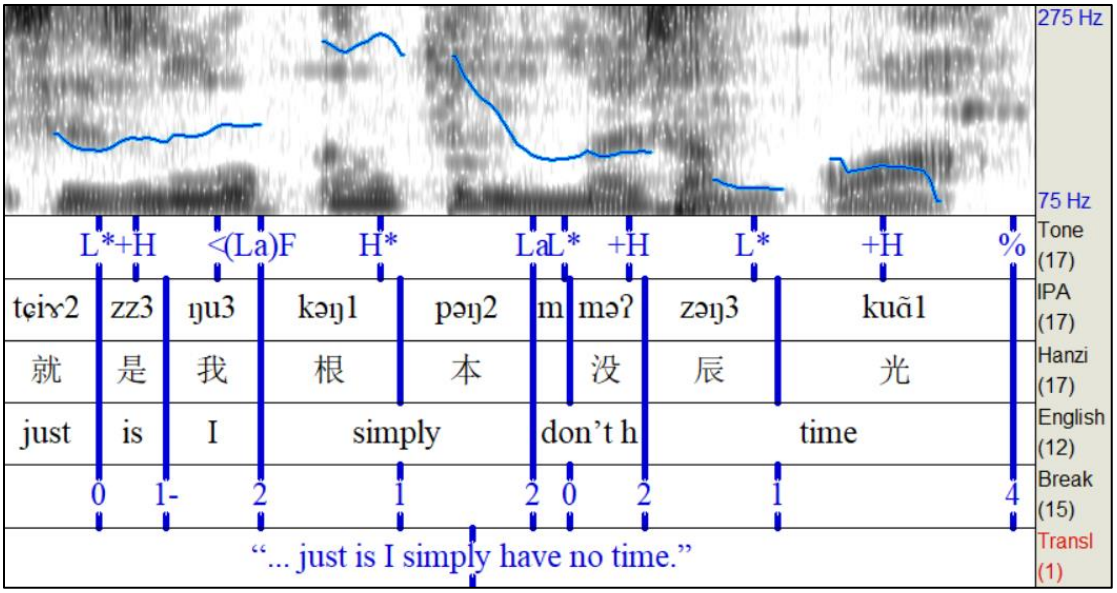

Figure 29. Annotated Shanghainese phrase '… just is I simply don't have time,' showing effects of focus on 根本 'simply;' in this case, a raised H* target, reduction of preceding AP 就是我 '…just is I…,' and pitch compression on the following APs 没有辰光 'don't have time.' Speaker: 9a.

In Fig. 29, we see similar behavior concerning the focused element, but the preceding AP 就是我 '…just is I…' is different. In Fig. 28, the speaker phrased the 就是 'just is' and 我 'I' as separate APs, lining up a prosodic boundary with the syntactic boundary before the embedded clause 'I simply have no time.' In Fig. 29, however, we see that this speaker has phrased the three syllables together into one AP, and the +H target is delayed to the point that no La is realizable at the end of the end of the last syllable. While both examples show the reducing effect that focus has on preceding APs, the utterance in Fig. 29 shows a more extreme application of reduction where a major syntactic boundary has been ignored in phrasing.

In both cases however, we can test the de-accenting predictions of S&S. Given their explanation of de-accenting and ip/major phrase determination, we should see de-accenting on 没有 'don't have,' as it is in the same VP as the focused element (an adverb), but does not

initiate its own lexical XP. As for 辰光 ‘time,’ a new NP is initiated, meaning for S&S that a new major phrase/ip should begin, and de-accenting should not affect the last two syllables. Comparing the predictions to the observed pitch tracks, we can see that neither speaker de-accented 没有 ‘don’t have’ in their utterance, though the speaker in Fig. 29 does seem to compress her pitch range a bit more than the speaker in Fig. 28. In the case of 辰光 ‘time,’ we see that the speaker in Fig. 28 apparently ‘de-accents’ the final syllable, though this more likely attributable to its IP-final position (see Section 4.7.2.3. on tonal reduction in such contexts). Regardless, we can see that de-accenting has failed to apply where S&S predict it should. Additionally, though they predict a phrase break after the focused major phrase that should be larger than an accentual phrase, there is no percept of juncture to native speakers or noticeable segmental lengthening, both of which could be expected for the boundary proposed here. Such examples of narrow focus without de-accenting are common in my data.

On the other hand, in the following Fig. 30, we can see another instance of focus (this time corrective) where de-accenting has over-applied according to S&S. In Sh_ToBI, such focus is marked by ‘aC’ preceding the focused AP. This sentence, “Dad DID’T LIKE mom’s gift,” is part of a longer, scripted dialogue where the speaker played both parts. In this situation, ‘Speaker A’ asks ‘Speaker B’ “Did dad like Mom’s gift?” The following sentence is the response given by ‘Speaker B.’

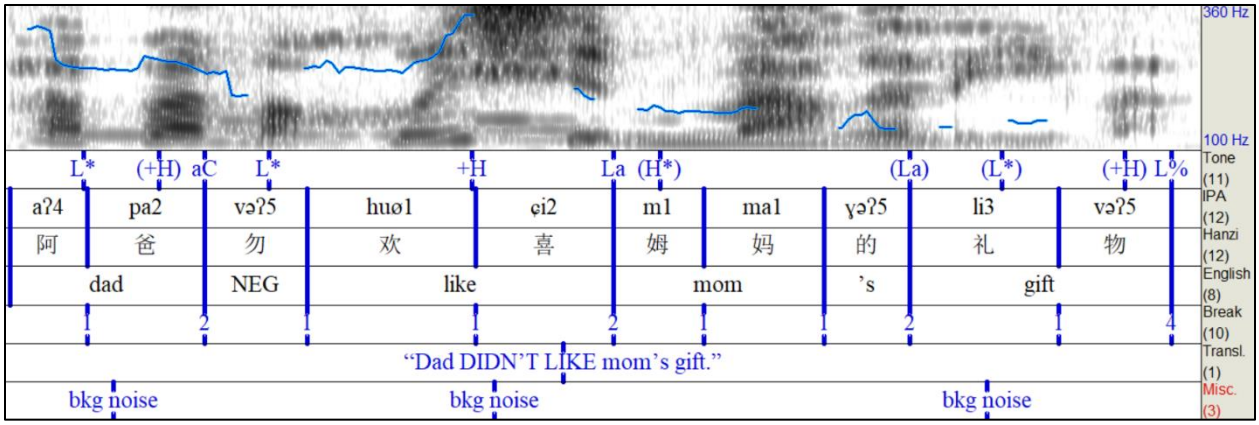

Figure 30. Annotated Shanghainese sentence 'Dad DIDN'T LIKE mom's gift' said in response to the question 'Did dad like mom's gift?' The pitch track shows focus on 勿欢喜 'didn't like' followed by complete de-accenting of all following APs. The most expected phrasing and tones on de-accented syllables is shown in parentheses. Speaker: 4a.

Again, a raised high target appears on the correctively focused AP 勿欢喜 'didn't like;' however, following this focused AP, all tones seem to be deleted, showing a generally low plateau to the end of the sentence with increasing creakiness. According to S&S, no de-accenting should occur at all: since 姆妈的 'mom's' initiates a new noun phrase, it is predicted to also initiate a major phrase, thus preventing de-accenting in following APs. Thus we can see that S&S's predictions about the intermediate phrase and its relation to focus do not match the data here in two ways: in Figs. 28 and 29, de-accenting 'under-applied,' while in Fig. 30 it 'over-applied.' This difference in post-focal de-accenting seems to be due to the type of focus, as all speakers had pitch range compression in narrow focus sentences (like in Figs. 28 and 29), while clearly corrective sentences (like Fig. 30) were consistently produced with full post-focal de-accenting by all speakers. Despite the categorical effect in my data, due to the relatively small number of corrective focus tokens, more work would need to be done to conclusively prove this behavior. Regardless, the fact remains that we cannot use the presence or absence of de-accenting as a diagnostic for finding an intermediate phrase as claimed by S&S.

4.6. The Intermediate Phrase

Instead of predicting deterministically from the syntax where intermediate phrase boundaries should be, longer, multi-clausal sentences were given to speakers in order to elicit different phrasing strategies. Working closely with a linguistically-trained native speaker, we identified places where she had a percept of a pause or a break larger than an AP, despite a lack of actual silence following the boundary. After identifying these contexts, there were two additional commonalities that seemed to unite the sentences with perceived pause. Together, the following four effects make up the diagnostics for annotating an intermediate phrase (ip) boundary: 1) percept of juncture larger than AP boundary; 2) lack of silence following said boundary; 3) segmental lengthening in ip-final syllable; 4) pitch reset after the boundary. In all cases of canonical ip boundaries, all four features must be present. Additionally, there may be contour reduction on the final AP of the ip despite segmental lengthening, though this is not found in all cases. In the remainder of this section, these identifiers for the ip will be discussed and compared with the behavior of the smaller AP boundary. Comparison with the largest unit of phrasing, the Intonational Phrase (IP) can be found in Section 4.7.2.3. In the data analyzed, ip boundaries were most commonly seen at clause boundaries, between topics and comments, and in some cases pre-focally; however, there is no definitive algorithm presented here for predicting their occurrence due to the variability of the data encountered.

In the following figures (Figs. 31-32), two different speakers use different phrasing strategies for the same sentence: ‘If he comes, Little Li will just want to leave.’ In a conditional sentence like this, the most expected place for an ip break to occur would be at the clause boundary—after ‘comes’ and before ‘Little Li.’ In these cases, one speaker (Fig. 31) has the entire sentence phrased in a single ip, with no boundary visible, while in the other (Fig. 32), the

speaker breaks up the sentence into two ips, with a clear boundary that has all features mentioned above. In Sh_ToBI, the boundary is annotated with a single lowercase **i**.

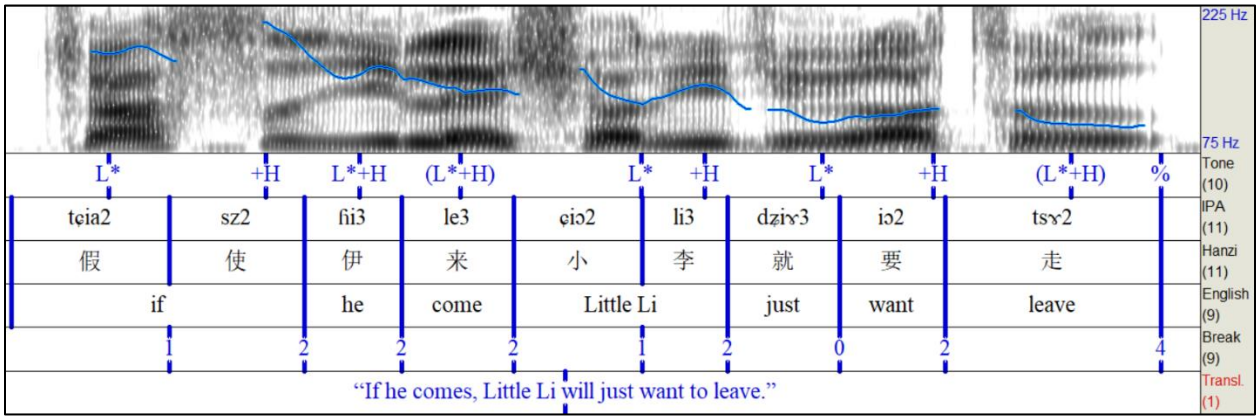

Figure 31. Annotated Shanghainese sentence 'If he comes, Little Li will just want to leave' phrased into a single ip. Speaker: 1c.

In Fig. 31, we see a long sentence parsed into a single ip, said at a fairly high speech rate. Though the monosyllabic verb 来 'come' has been reduced, we can see no lengthening of its segmental material (compared with Fig. 32 below). Additionally, looking at the pitch across the clause boundary we see a normal falling relationship between the flattened pitch on 来 'comes' and the lower L* target that follows on 小李 'Little Li.' In general, the pitch range gets smaller over the course of the sentence, to the point that the last AP has been reduced to a gradually falling tone, regular for the end of a long IP. For the native speaker consultant, there is no percept of a juncture larger than an AP at the clause break. Thus, this utterance behaves as the others we have seen up until this point, as each has consisted of a single ip.

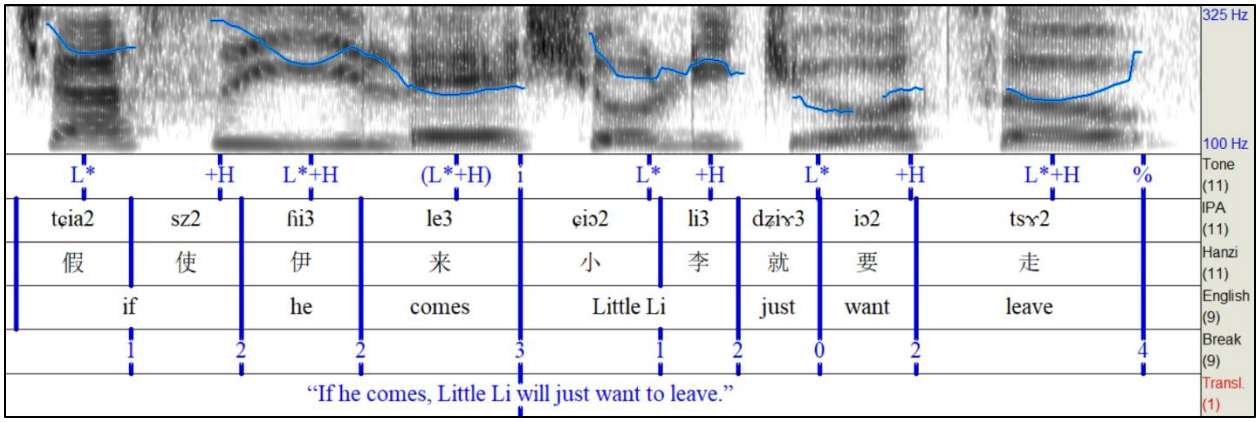

Figure 32. Annotated Shanghainese sentence 'If he comes, Little Li will just want to leave,' showing an ip boundary between the two clauses. Speaker: 7c.

In Fig. 32, however, the intermediate phrase boundary lines up with the large syntactic break between the two clauses. This is evident in the lengthening of the ip-final syllable, 来 'come,' and the lack of silence following it. Additionally, we see that the following L* on 小李 'Little Li' is actually *higher* than the reduced tone on 来 'come.' Since we know that L* should generally be a local pitch minimum, we can tell that pitch reset has also applied over this boundary. These factors together create the percept of a juncture larger than an AP, but smaller than an IP for my native speaker consultant, and thus can be used to clearly identify ip boundaries wherever they occur, regardless of expectations based on syntactic structure.

In this section, Selkirk & Shen's claims about focus and their major phrase (here, the intermediate phrase) were found not to hold for the speakers examined. Firstly, their predictions for post-focal de-accenting were not supported, as speakers only de-accented in corrective focus sentences, not emphatic sentences. In the cases where de-accenting did apply, it did not follow their predictions about only applying within the ip containing the focused AP, and instead applied to all post-focal APs in an IP. Additionally, their deterministic algorithm for determining ip boundaries was untenable for the data collected here, as speakers do not have ip breaks at the

beginning of every lexical XP. While ip breaks are most common at large clause boundaries, there is still a fair amount of variation that makes a strong generalization premature at best. Finally, four actual phonetic diagnostics for the ip were identified: 1) percept of a juncture larger than an AP but smaller than an IP; 2) lack of following silence; 3) segmental lengthening before the boundary; and 4) pitch reset following the boundary.

4.7. The Intonational Phrase

The Intonational Phrase (IP) is the highest level of phrasing in Shanghainese, which generally ranges in size from a large clause to entire utterance, and must contain at least one intermediate phrase (ip). The IP has three main identifiers: optional initial pitch range expansion, and obligatory final boundary tones and final lengthening. IP-initial pitch range expansion was first described in my master's thesis (2016); it is a process where the upper limit of the pitch range is raised during the first accentual phrase (APs) of the IP, resembling focus without causing any de-accenting or pitch range compression on following APs. The analysis of IP-initial pitch range expansion (IPRE) has not changed significantly between my previous analysis and the one presented here. Below, the behavior and application of IPRE will be discussed, showing its interactions with the prominence imbalance between H* and +H tones. Another marker of the IP is the final boundary tones, all of which are always accompanied by significant segmental lengthening and a following silence. In the system proposed here, there are three contrastive boundary tones, each with their own meanings and phonetic realizations. H%, with a high target, is used in information-seeking questions and interacts with other tones variably; L%, realized with a low tone that crowds onto IP-final syllables, is used for correction, frustration, or emphasis; lastly, % is the most common boundary tone, used in unmarked statements and many

kinds of questions. While it lacks a distinct pitch target,[15] % induces IP-final pitch range compression and lowering, often with full tonal reduction of the final syllable.

4.7.1. IP-Initial Pitch Range Expansion

IP-initial pitch range expansion (IPRE) is a process that ensures the highest pitch of a Shanghainese utterance is in the first AP, and resembles focus in many ways. The upper limit of the pitch range is always raised, and the lower limit can also be lowered. Pitch peak delay can also occur. While not present in all IPs in the data, it is not yet clear what determines its presence or absence. Because it is so pervasive and generally predictable, there is no distinct annotation for its occurrence in Sh_ToBI; however, if there is a conspicuous lack of IPRE where expected, it could be noted by the annotator in the Misc. tier. The most common context for a lack of IPRE is at the beginning of a subordinate clause preceded by a silence. Fig. 33 below shows a normal example of IPRE:

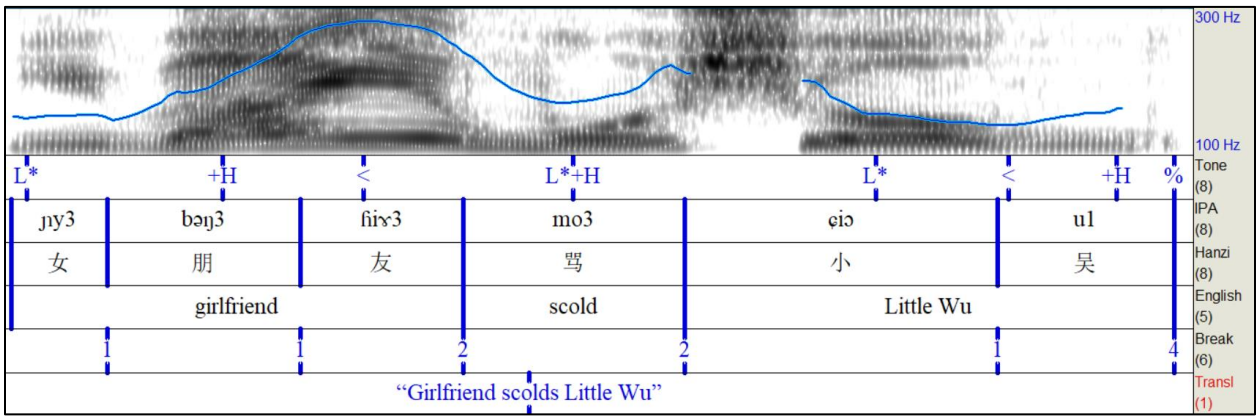

Figure 33. Annotated Shanghainese sentence 'Girlfriend scolds Little Wu,' showing the f0 maximum in the first AP due to IP-initial pitch range expansion. Speaker: 6c.

[15] Note that while uncommon, toneless boundary 'tones' have been reported in other languages, both lexically tonal and not, like Cantonese (Wong et al. 2005) and German (Grice et al. 2005). In these analyses, % represents an IP-level of lengthening and juncture, but without tonal information is attributable to the IP-level of phrasing. In these cases, the last specified tone in the IP is left unaltered (Cantonese) or plateaued until the end of the IP (German).

While IPRE generally only affects the first AP of an IP, the second AP can sometimes have the pitch peak without focus. This can happen in one of two ways, only one of which will be discussed in this section (see Section 5.3.1. on Break 2m for discussion of the other). In the relevant case for this section, the prominence relationship between an IP-initial AP's +H and the second AP's H* can override IPRE to ensure that H* remains higher than the +H target on the first AP; however, this is not always the case, as can be seen in comparing the following two figures (Figs. 34-35), where the same speaker says two +H)(H* sequences, taking different approaches to where the pitch peak of the IPRE should be.

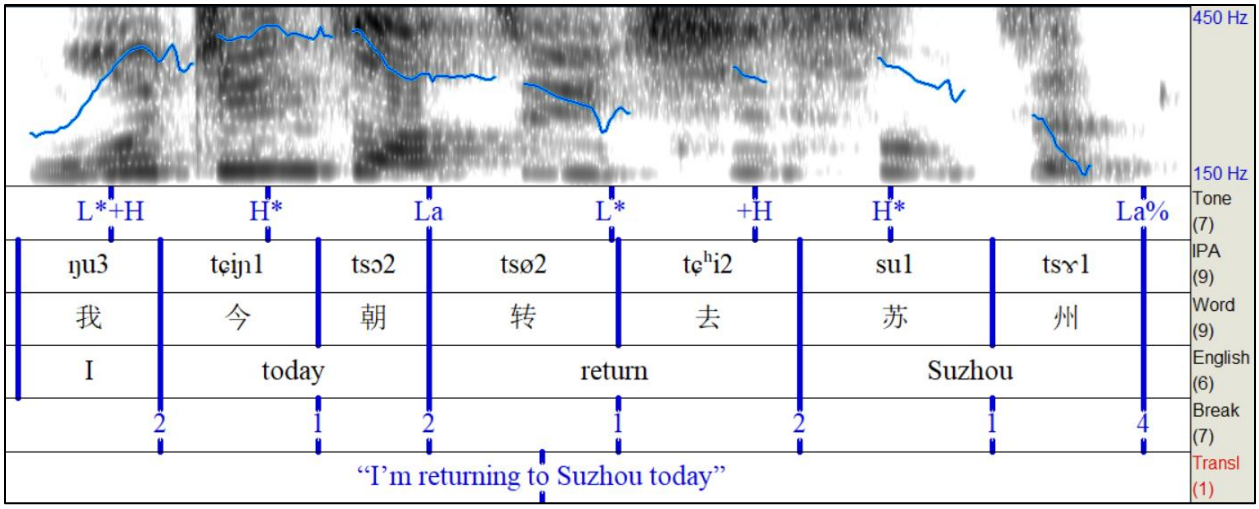

Figure 34. Annotated Shanghainese sentence 'I'm returning to Suzhou today,' showing the prominence imbalance between +H and H* overriding IP-initial pitch range expansion. Speaker: 1a.

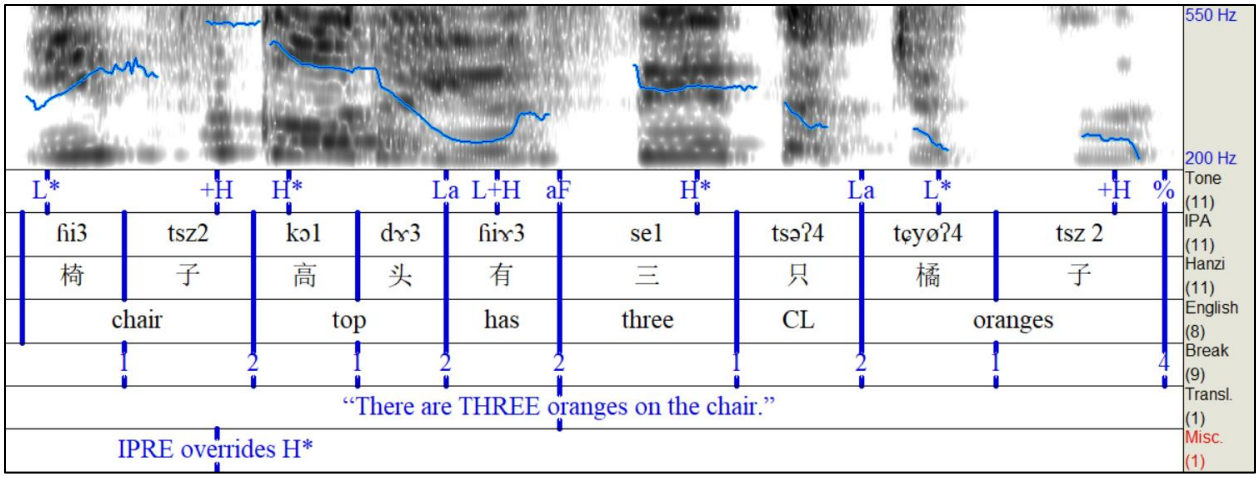

Figure 35. Annotated Shanghainese sentence 'There are three oranges on the chair,' showing IP-initial pitch range expansion overriding the prominence imbalance of +H and H*. Speaker: 1a.

In Fig. 34, we see a case where the prominence imbalance between the high tones has won out, we see the familiar pattern the H* on the second AP 今朝 'today' is higher than the +H in the first AP 我 'I.' In Fig. 35, IPRE overrides the prominence relationship, and the IP's pitch peak is found on the first AP, 椅子 'chair.' While there is a syntactic difference between the AP boundaries here (the first spans the subject-VP break, while the second is subject-internal), there was no consistency across speakers along this line; instead it may have to do with the phonological weight of the first AP—speakers seem to prefer the second AP's H* as the highest pitch when the first AP is monosyllabic and/or non-lexical, but the first AP's +H is highest when the first AP is disyllabic; however, more work needs to be done to confirm this.

4.7.2. IP-final Boundary Tones

IP-final boundary tones are present at the end of every IP in Shanghainese. In my previous work, I proposed four contrastive IP boundary tones: L%, M%, H%, and ^H%. The proposal here only has three contrastive categories, toneless %, H%, and L. Originally L% was the basic, unmarked statement IP-boundary tone, but with the discovery of a real L-target bearing L%, the unmarked tone was recognized to be toneless %, realized via segmental lengthening, pitch range compression, and optional tonal reduction. M% was originally used for continuation or listing, but was subsumed under % when it was realized they were not contrastive to my native speaker consultant. ^H% was also found to be a combination of focus occurring on IP-final APs and H%, and thus the system here does not include it. In the following table, the three current IP-boundary tones are listed with their realizations and generalized meanings.

IP Boundary Tones	Realization	Meaning
H%	H tone target at end of IP with potential f0 range expansion/raising towards end. Interacts variably with preceding L targets	Particle-final questions; echo questions; disbelief
L%	Strong, separate L target; will form a clear contour on IP-final H tone syllables	Marked statements; negative, corrective, emphatic, obviousness
%	No clear separate pitch target, only realized via phrase-final lengthening, pitch range compression, and optional tonal reduction	Unmarked/general statements; wh-questions; A-not-A questions; particle-final questions; rhetorical questions

Table 10. Inventory of IP-final boundary tones with their realizations and meanings.

4.7.2.1. High IP Boundary Tone – H%

Like many languages, Shanghainese uses a high IP-final boundary tone, H%, to mark questions. Before continuing, it is important to distinguish the four types of questions investigated here: particle yes/no, A-not-A yes/no, wh-questions, and rhetorical questions. In most Chinese languages, there are generally two basic ways of forming yes/no questions. The first involves a particle, e.g. 伐 *va* in Shanghainese, that is inserted at the end of a sentence to indicate a question. The other, A-not-A questions, involve a reduplication of the main sentential verb with the negative particle, 勿 *vəʔ* in Shanghainese, inserted between the two copies of the verb; again this is interpreted as a simple yes/no question, with little to no semantic or pragmatic difference from the particle yes/no questions. Wh-questions are those which involve some sort of wh-element (e.g. who, what, where, etc.; formed with 啥 *sa* 'what' in Shanghainese) and expect an answer consisting of some lexical phrase; like many Chinese languages, wh-elements remain in-situ in Shanghainese. Lastly, rhetorical questions are not information-seeking, but rather a device to confirm or suggest a proposition. In the data analyzed for this dissertation, H% was only found to occur for some particle-final yes/no questions—all others consistently used the toneless boundary, %, instead.

In terms of tonal realization, H% interacts any La tone(s) associated to the IP-final syllable by either combining with it (Fig. 36) or deleting it outright (Fig. 37). However, in quadrisyllabic IP-final APs, La moves earlier to accommodate the H% (Fig. 38). These behaviors are explained below. As with all IP-final boundary tones, it is accompanied by significant segmental lengthening and a following silence.

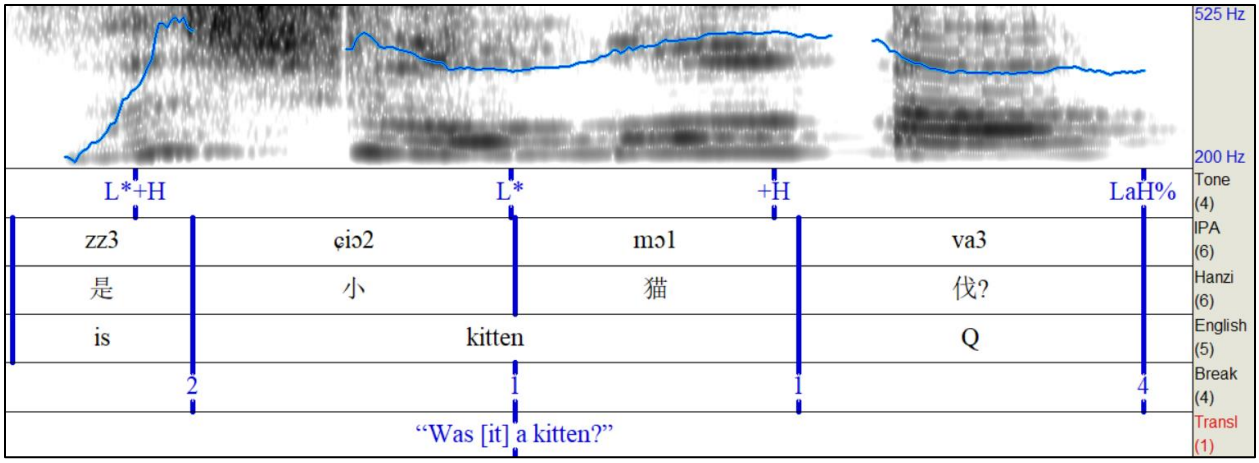

Figure 36. Annotated Shanghainese sentence "Was it a kitten?" showing the use of H% in a particle yes/no question. Here, the final La has been overridden and replaced with a moderately high level tone. Compare with Fig. 48 below. Speaker: 1a.

In this first example of H%, the speaker is playing both parts of dialogue, where the first 'speaker' tells the second that they saw a puppy yesterday. In response, the second 'speaker' in the dialogue says they didn't hear clearly, and asks "Was it a kitten?" Instead of a normal falling pitch over the last syllable to the La (see Fig. 48 below), the AP-final La has combined with the H% target, creating a level tone at nearly the same pitch as the earlier, high-variant L* target on 小猫 'kitten.' Despite being in line with a low tone target, the pitch remains much higher than expected for a % boundary, as seen in Fig. 48 below. This process is characterized as 'combination' due to the existence of other strategies.

For instance, in Fig. 37, we see an H% that has completely overridden the La that should appear on the final syllable of the IP.

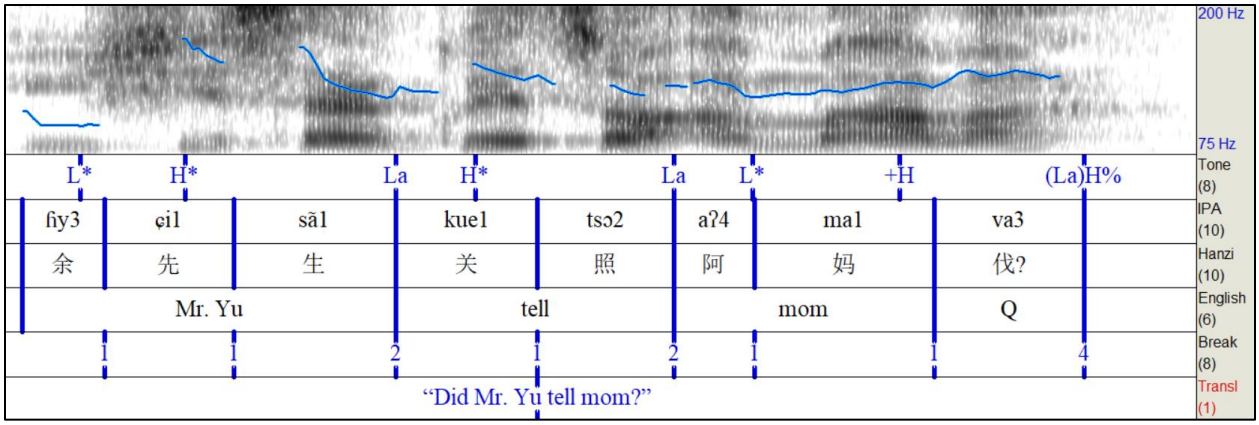

Figure 37. Annotated Shanghainese sentence "Did Mr. Yu tell mom?" showing an H% in a question-particle-final sentence. The IP-final La is completely overridden by a steady rise against declination, ending higher than the immediately previous +H. Speaker: 10a.

In this utterance, instead of creating a level tone lower than preceding H targets, the H% is actually higher than the previous +H, making for a gradual rise to the end of the IP, showing the deletion of the La from the surface pitch realization. At this point, it is still unclear why speakers use one strategy (combining La and H%) over another (deletion of La). While all speakers that produced a H% in the sentence in Fig. 36 ('Was it a kitten?') produced the combined level tone, speakers differed in what strategy they used in the sentence in Fig. 37 (Did Mr. Yu tell mom?). More work would need to be done to see if this is simply an effect of IP length or if this is due to some dialectal or age-based differences.

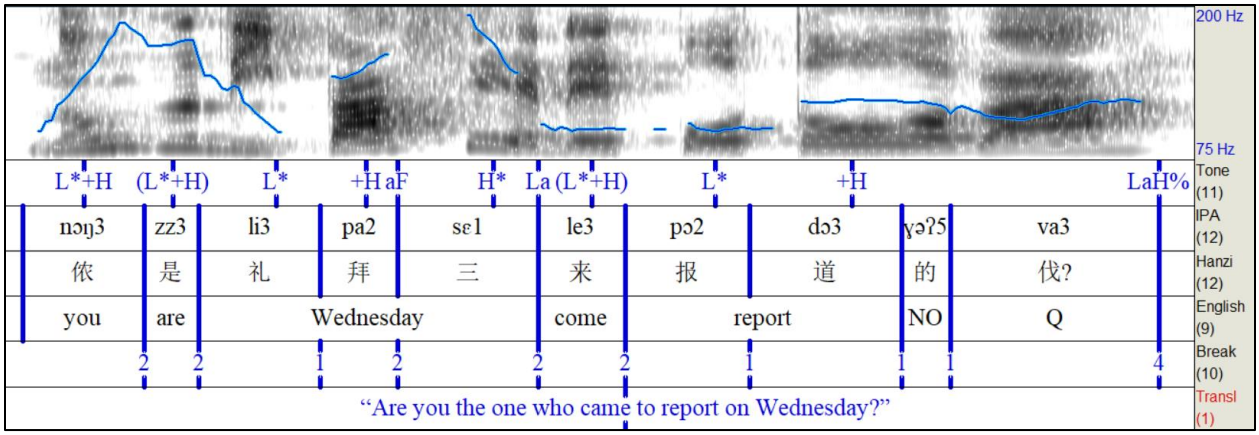

Figure 38. Annotated Shanghainese sentence "Are you the one who came to report on Wednesday?" showing the use of H% in a particle yes/no question. IP-final AP is quadrisyllabic, with enough space for both the La and H% to be realized, creating a rise over the last syllable. Speaker: 5a.

In this last example of H% (Fig. 38), we see a sentence with a long, quadrisyllabic AP, 报道的伐 '[the one who] reported?' at the end of a question. Instead of the H% overriding the final La, the La target is pushed back to accommodate both tones, leading to a small rise over the last syllable. This pattern was observed with all speakers, regardless of their strategy for reconciling La and H% in trisyllabic APs (i.e. combining La and H% or deleting La). This pitch track can also confirm the focal effect seen in Tian & Kuang (2020), where they show that speakers force a new AP boundary when a non-compound-initial syllable/morpheme is focused. In Fig. 38, the word 'Wednesday' has been separated broken up to focus the final of the three syllables. In Chinese, the days of the week are numbered, and Wednesday is 礼拜 'week' 三 'three'—here the disambiguating morpheme, 'three' has been focused and thus phrased as its own AP.

4.7.2.2. Low IP Boundary Tone – L%

The low boundary tone, annotated L%, is used in negative, corrective, or otherwise emotionally-marked statements. It is interesting to note that several speakers used this boundary tone, despite the fact that there were only a few stimuli that had enough context to prime the speakers for the use of L%. Because of the lack of stimuli specifically constructed to elicit this tone, examples of L% in the data here are fairly rare. Thus, L% has only been identified in IPs whose ends coincide with either +H or LHa. Unfortunately, at the time of writing, there is no available data that can show the effect of L% on either +L or La. In the examples that follow, we see that L% always forms a contour with the preceding H target, never overriding tones as H% does. Thus it appears that Shanghainese's strategy in handling tonal crowding is dependent on the identity of the intonational tone itself.

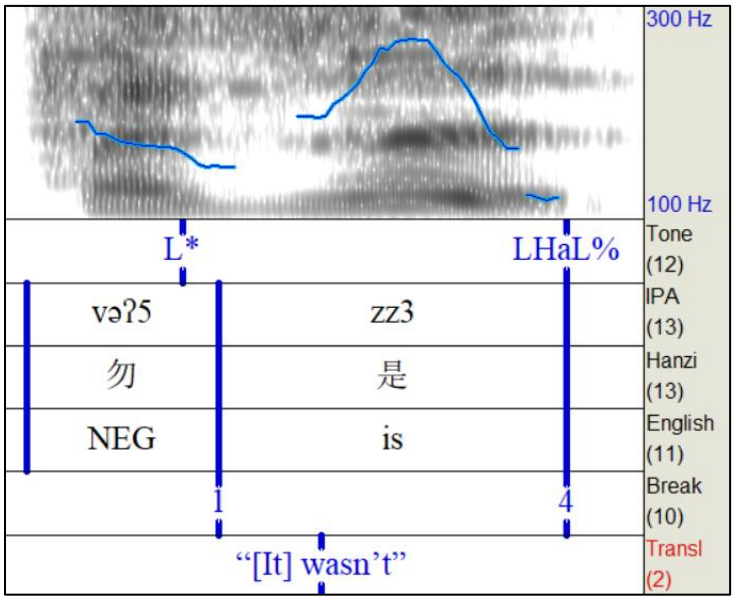

Figure 39. Annotated Shanghainese sentence 'It wasn't' as part of a dialogue where the speaker was pretending to correct someone. The final L% does not override the AP-final boundary tone LHa. Instead, it creates a sharp rising-falling pattern over the final syllable. Speaker: 4a.

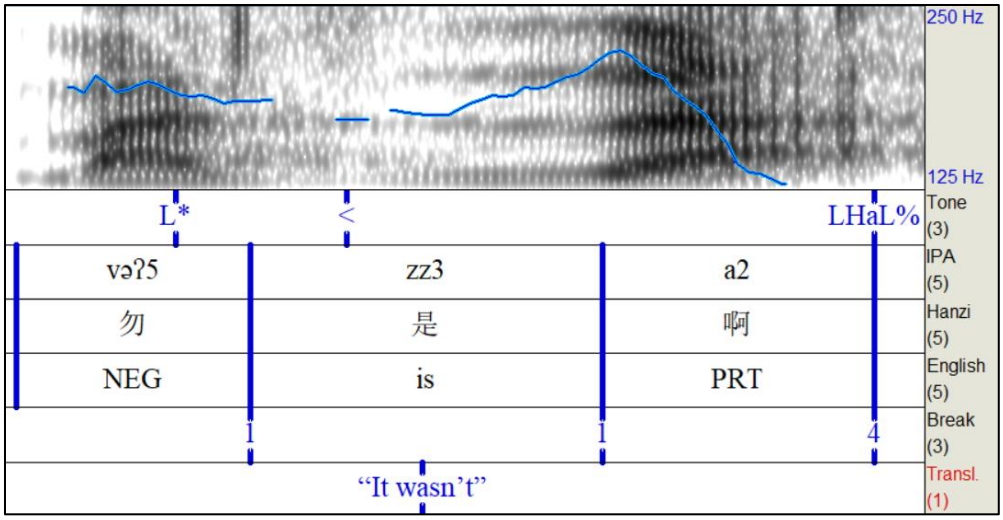

Figure 40. Annotated Shanghainese sentence "It wasn't." as part of a dialogue with the speaker pretending to correct someone. The final L% does not override the AP-final LHa phrase accent, but does push it back onto the penultimate syllable of the AP, with a sharp fall on the last syllable. Speaker: 5a.

In these two examples (Figs. 39 & 40), we see a case where context clearly guides the use of L%. The speaker plays both parts in a dialogue where 'first speaker' asks whether the 'second speaker' saw a cat; the 'second speaker,' in response, corrects them, using an L% at the end of the short IP 勿是 'No, it wasn't [a cat].' In Fig. 39, the AP is only two syllables long, so the L% squeezes onto the IP-final syllable which also plays host to the AP-final LHa tone, resulting in a steep rising-falling contour; this seems to be the maximal number of tonal targets that can be realized on a single syllable. In Fig. 40, however, we see that the speaker added a final modal particle, potentially to add emphasis or a sense of 'obviousness' to the sentence, but also perhaps in part to a dispreference for three tonal targets on a single syllable. In this case, the LHa is actually pushed back a full syllable, being realized on 是 'is,' while the L% causes a steep fall over the final syllable. This can be compared with Fig. 41 below, where another speaker has used toneless % at the end of the same sentence, allowing for the LHa to be realized at the end of its AP normally.

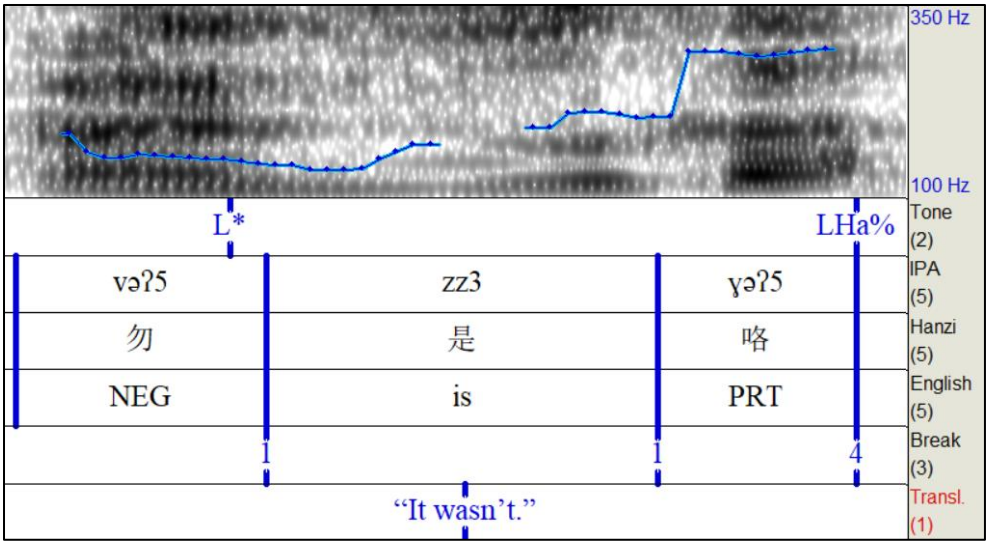

Figure 41. Annotated Shanghainese sentence 'It wasn't,' using toneless % instead of L%, showing normal LHa realization on the final syllable. Speaker: 4a.

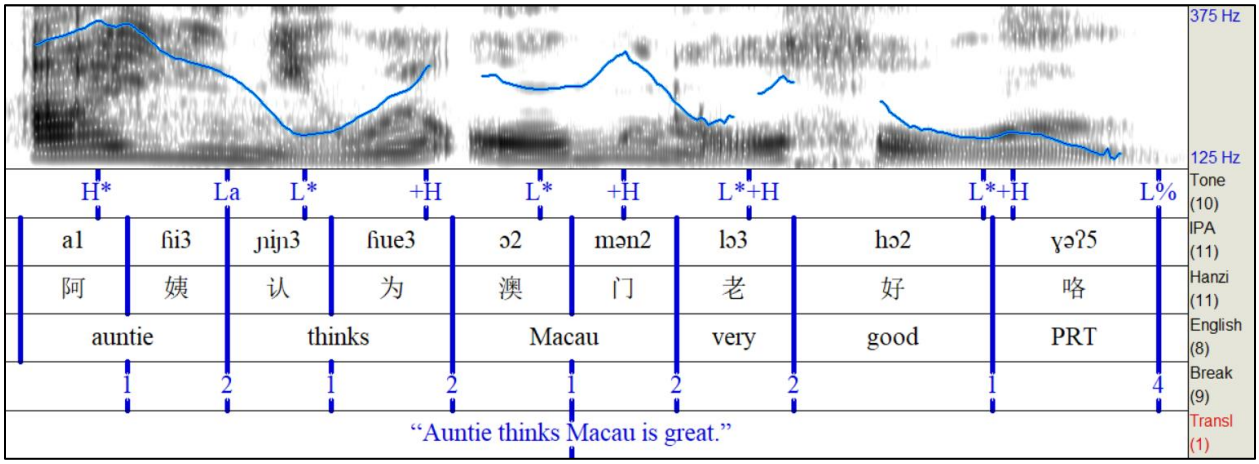

Figure 42. Annotated Shanghainese sentence "Auntie thinks Macau is great," showing an IP-final L% in an emphatic statement. The final syllable's +H target has been moved early, followed by a clear fall. Speaker: 1b.

In Fig. 42, the speaker decided to use an L%, perhaps to emphasize just how much 'Auntie thinks Macau [a city] is great,' despite the fact that there was no context given to this statement. Again, it occurs with a sentence-final modal particle. In this case, the L% must co-exist with a +H target on the last syllable. As with the previous examples, the +H target is moved

earlier in its syllable (though not on to the preceding syllable, as it already hosts L*), and the L% is realized afterwards, resulting in a falling pitch across the last syllable. While this fall is not nearly as sharp or dramatic as in the simple examples given above in Figs. 39 and 40, it is quite perceptually salient.

Lastly it appears that L% also occurs at the end of sentences with corrective focus (aC). This can explain the low plateau pattern seen in de-accented syllables post-focally, and it also fits the pragmatic use of L% in other cases. At this time, it does not appear that corrective focus can occur with any other IP-boundary tone. Fig. 43 below (repeated from Fig. 30) shows an example of corrective focus, with paired aC and L%.

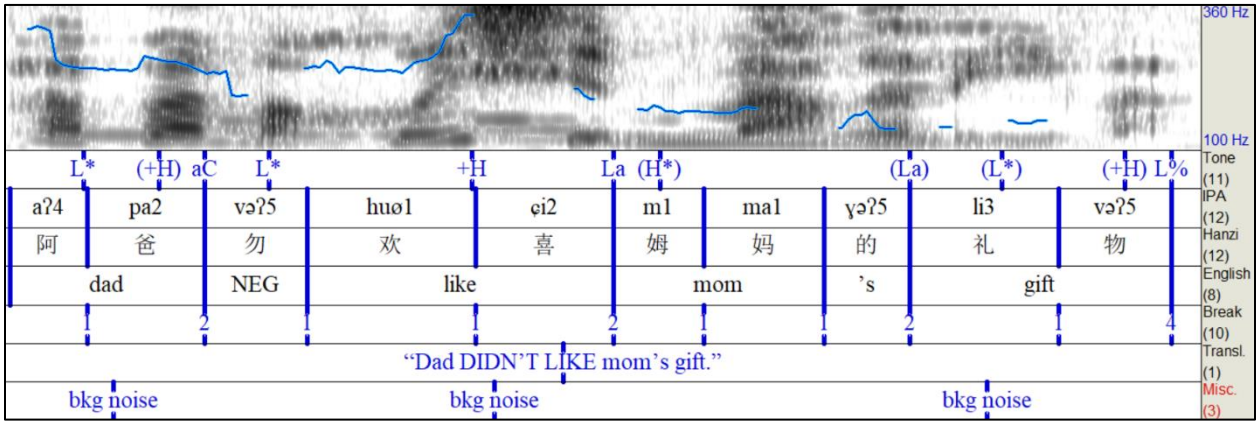

Figure 43. Annotated Shanghainese sentence 'Dad DIDN'T LIKE mom's gift,' with corrective focus, annotated aC, on the second AP 勿欢喜 'didn't like.' This is paired with the corrective IP-final boundary tone L%, creating a low, de-accented plateau. Same as Fig. 30 above. Speaker: 4a.

4.7.2.3. Unmarked Uses of % – the Toneless IP Boundary 'Tone'

The toneless IP-boundary 'tone' is by far the most commonly used IP boundary in Shanghainese, being used for almost all general, unmarked statements. Beyond this, it can be used in other contexts (examples following below), such as questions and clause boundaries in longer utterances.

Despite not having a discernable pitch target, % is still evident through two main phonetic effects: phrase-final lengthening of the IP-final syllable and a following silence. Additionally pitch range compression, tonal reduction, and creakiness can variably accompany % as well. Consider the following two sentences (Figs. 44 and 45) produced by the same speaker, paying attention to the final, monosyllabic AP:

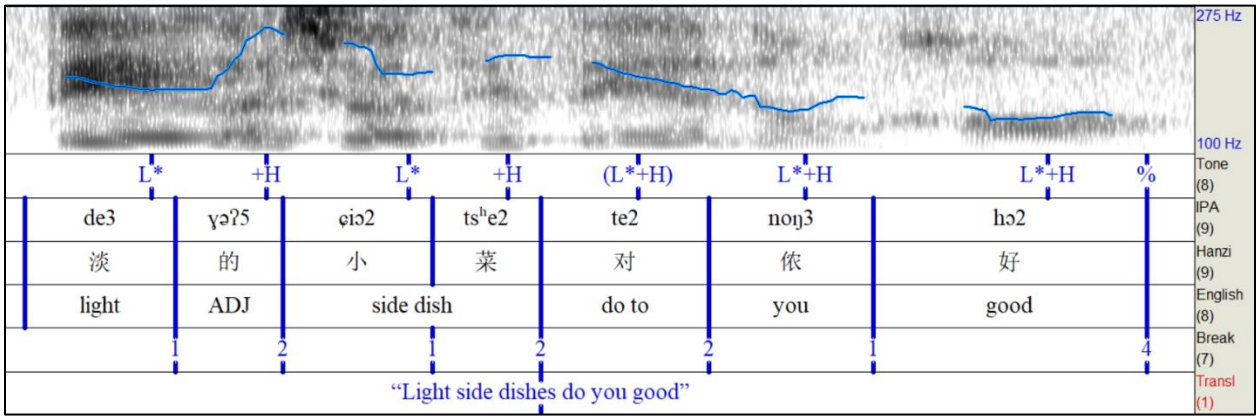

Figure 44. Annotated Shanghainese sentence 'Light side dishes do you good,' showing a small but clear rise on the IP-final L*+H AP, despite a following %. Speaker: 15c.

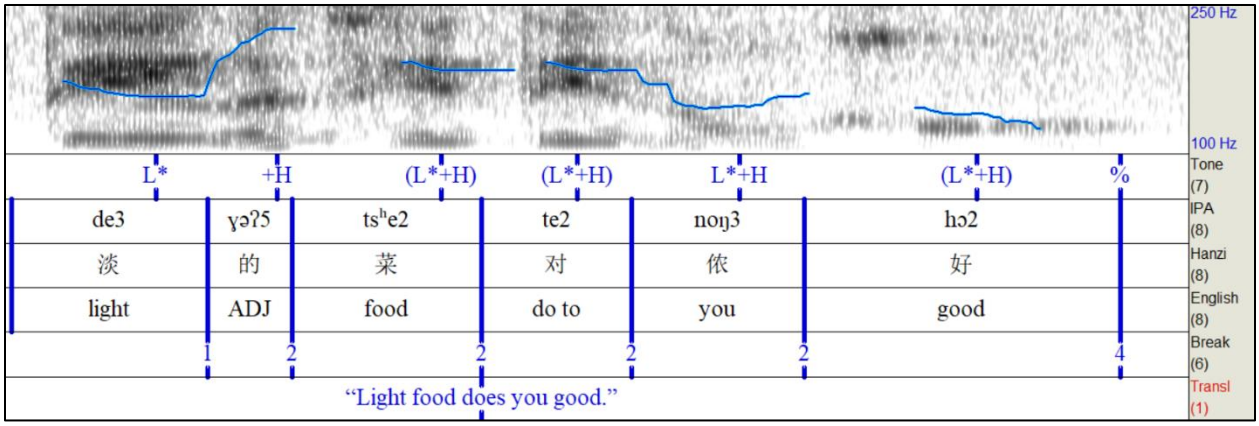

Figure 45. Annotated Shanghainese sentence 'Light food does you good,' showing a completely reduced tone on the IP-final L*+H AP, due to a following %. Speaker: 15c.

In the Fig. 44, the L*+H contour is realized as expected, though with reduced pitch range. However, in Fig. 45, the L*+H contour is replaced with a low falling tone, showing complete reduction of the tone (similar to the 对 'do to' in Fig. 44, which would be rising if not reduced). While one may expect Fig. 45 to be labeled as L%, it does not meet the phonetic or pragmatic requirements laid out below for such a boundary tone. Additionally, for my native speaker consultant, there does not seem to be a contrastive difference between these two sentences pragmatic meaning, and at first listen, she did not notice a difference in pronunciation between the two at all. Thus, the IP-final phenomena in Figs. 44 and 45 have been treated as the same category.

In both of these examples, % has been used at the end of unmarked, general statements. However, in the following examples (Figs. 46 and 47), we will see that % can be used in multi-clausal sentences and questions as well. In some cases, rather than an ip boundary, i, a speaker will use a full IP boundary to break up a long multi-clausal statement. Compare the two nearly identical sentences as produced by different speakers:

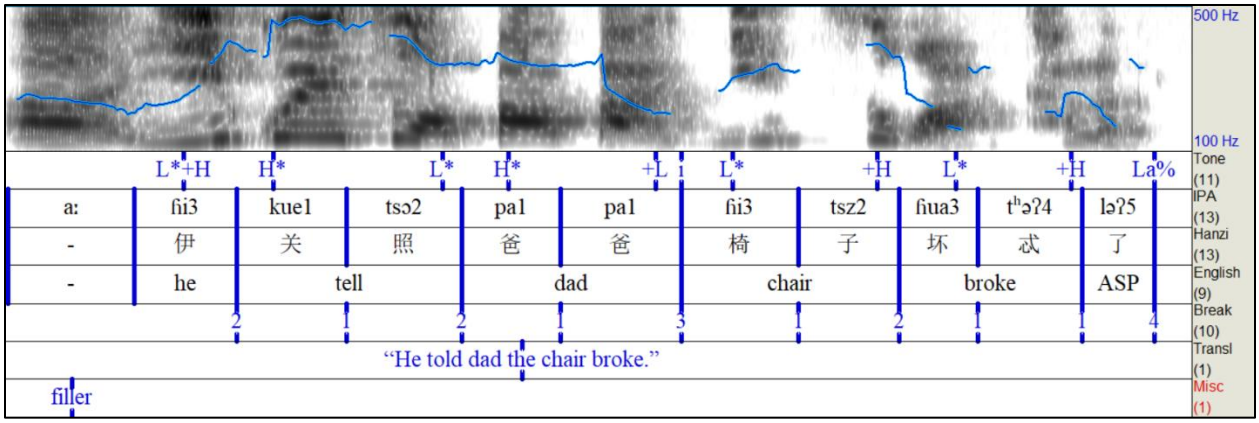

Figure 46. Annotated Shanghainese sentence 'He told dad the chair broke,' showing an ip boundary at the clause boundary between 爸爸 'dad' and 椅子 'chair.' Speaker: 1a.

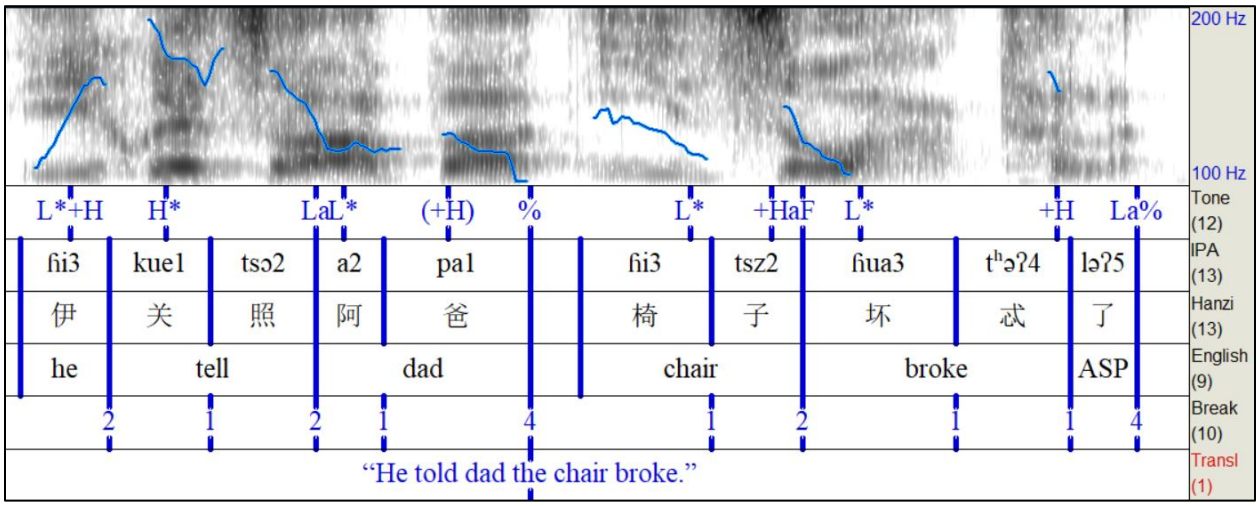

Figure 47. Annotated Shanghainese sentence 'He told dad the chair broke,' showing an IP[16] boundary at the clause boundary between 阿爸 'dad' and 椅子 'chair.' Speaker: 10a.

The first speaker (Fig. 46) has used a smaller ip boundary, i, to break up the two clauses, while keeping the whole sentence in a single IP. Conversely, the second speaker (Fig. 47) has used a larger IP boundary, %, breaking the sentence up into two full IPs. Both % and i are relized

[16] Note that second AP here may lack noticeable IPRE due to either or both: 1) the influence of focus on 坏忒 'broke;' 2) the fact that it is a subordinate clause and/or the second IP in an utterance.

with lengthening and potential tonal reduction; however, here the presence of silence is key. Because of the inter-clause silence, we can tell that latter speaker has used a full % boundary compared to the former speaker's shorter i boundary.

4.7.2.4. Use of % in Questions

The speakers asked to say questions (i.e. those from Dataset I) used % in every kind of question, and in fact, A-not-A questions, wh-questions, and rhetorical questions were only ever produced with %, never with H%. As already stated above, only particle yes/no questions were found to co-occur with H%, and even then some speakers still used % in these contexts. In the following examples, the use of % in each type of question is identified and explained.

In Fig. 48, a different speaker is acting in the same 'dialogue' as in Fig. 36 above. Despite the very clear information-seeking context provided, the speaker has used a basic %, which does not interfere with the realization of the AP-final La.

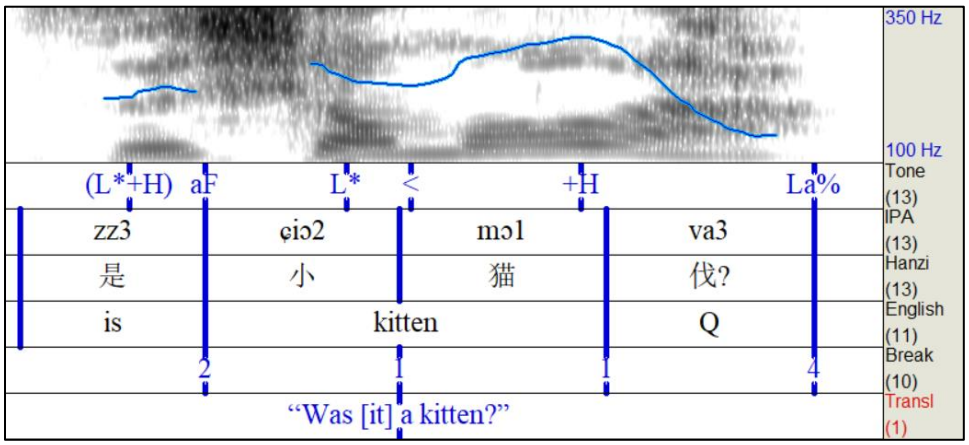

Figure 48. Annotated Shanghainese sentence 'Was it a kitten?' as part of a fake dialogue where the speaker was pretending to mishear someone. Despite being an information-seeking question, a normal % is used. Compare with Fig. 36 above. Speaker: 4a.

In Fig. 49, we see an A-not-A construction, which seems to be accompanied by a combination of prosodic cues: a % boundary tone and narrow focusing of the A-not-A complex. In this case, it also happens to be preceded by an ip break, though this aspect is not consistent across speakers. Again, this sentence is part of a longer dialogue where the speaker plays both parts, showing that the speaker most likely did intend a real question here.

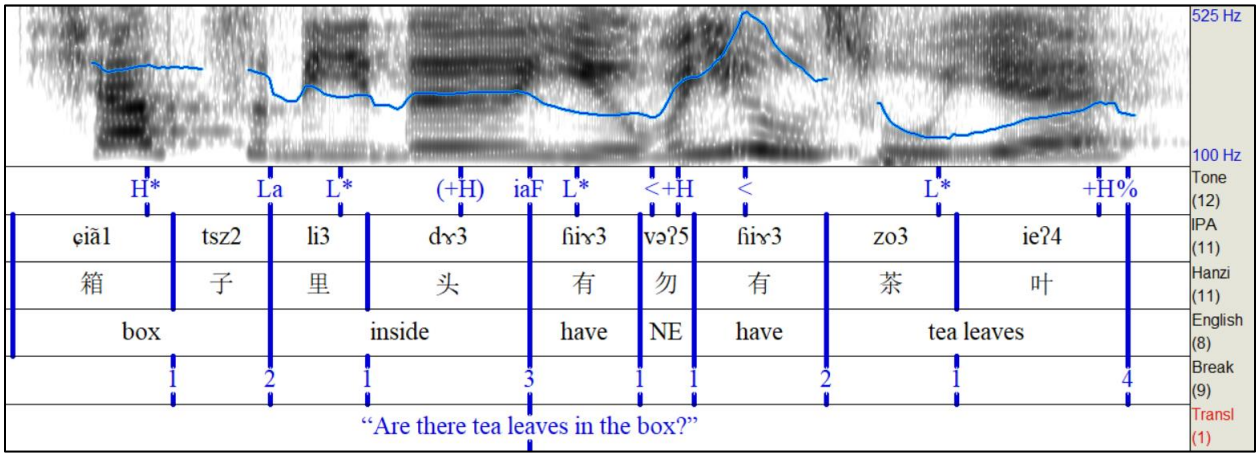

Figure 49. Annotated Shanghainese sentence 'Are there tea leaves in the box?' showing A-not-A question syntax. The A-not-A sequence is preceded by an ip break and is also focused itself. A normal % is used IP-finally. Speaker: 5a.

In Fig. 50, we see a wh-question, 'Do you know what Mrs. Yu is doing today?'[17] Instead of the tonal reduction normally associated with IP-final syllables, the speaker has produced an exaggerated L*+H contour on the wh-element, with the IP-final +H target realized as higher than the two preceding H targets. This in addition to the fact that the speaker has produced the final

[17] Note that while the idiomatic English translation is a yes/no question due to syntactic constraints, many Chinese dialects do not impose island restrictions on argument wh-elements (Huang 1982 for Mandarin), as is apparently the case here. Perhaps a closer, if more cumbersome, translation would be 'What is it that you know that Mrs. Yu is doing today?' Also note that while many speakers felt comfortable rejecting certain stimuli for unnaturalness, this sentence was always accepted.

AP with considerable lengthening of segmental material indicates 啥 ‘what’ has been focused (as well as being phrase-final).

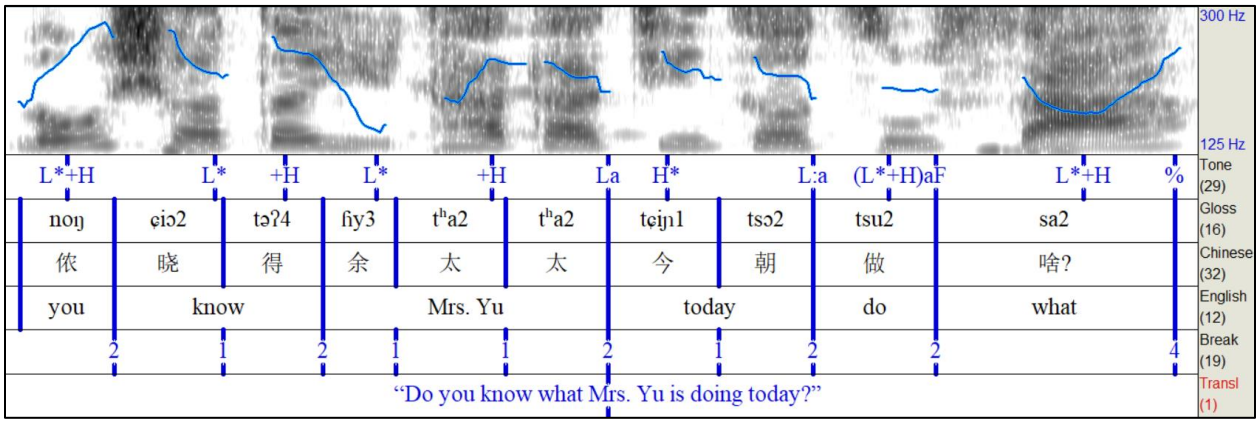

Figure 50. Annotated Shanghainese sentence “Do you know what Mrs. Yu is doing today?” showing focus on the final AP, the wh-element 啥 ‘what.’ Speaker: 4a.

The same strategy can be seen in Fig. 51 below, where the disyllabic wh-element, 为啥 ‘why’ appears sentence-medially. It is focused and followed pitch range compression. As wh-questions are not corrective, this behavior fits with the predictions made in the previous sections.

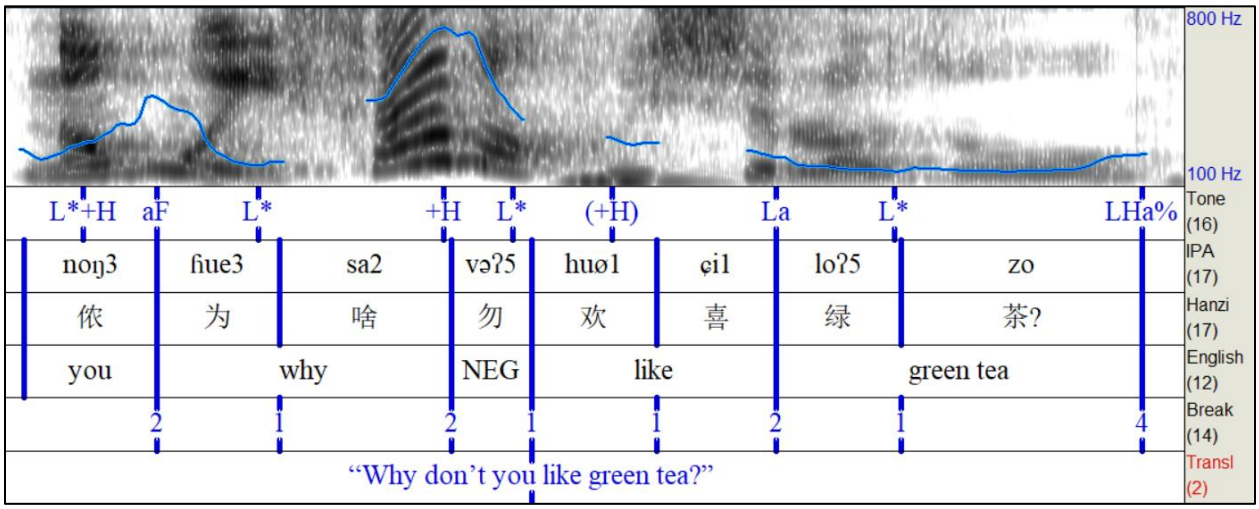

Figure 51. Annotated Shanghainese sentence ‘Why don’t you like green tea?’ as part of a fake dialogue, where the speaker is responding with shock to someone’s opinion. Here, the wh-element ‘why’ is sentence-medial, focused, and paired with a normal IP-final %. Speaker: 5a.

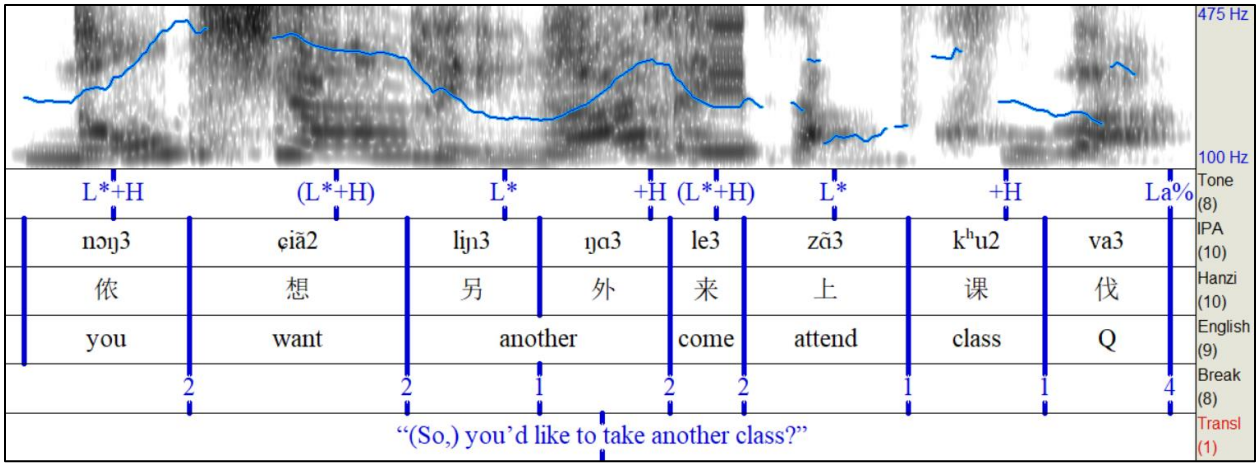

Figure 52. Annotated Shanghainese sentence 'You'd like to take another class?' showing the use of % as IP-final rhetorical question marker. Speaker: 1a.

In this last case, Fig. 52, we can imagine a scene where a student has come to a counselor of some sort who knows that the student wants to add a new class to their schedule. When the student comes in, the speaker's question is not really asking for the student's intentions; instead, they are looking for confirmation of known or obvious facts—they do not expect a no. Thus, the intonation seen here does not differ from that of the average statement.

4.7.2.5. Summary of IP Boundary Tones

In this section, the three contrastive IP boundary tones of Shanghainese were investigated: H%, L%, and toneless %. H% is used in particle yes/no questions, and can combine with, delete, or displace preceding low targets. L% is used in corrective, negative, or otherwise emotionally marked statements, often co-occurring with modal particles. It always crowds with or displaces previous tonal targets, never deleting or combining with them. Lastly, % is used in general unmarked statements, and in the vast majority of questions. Table 10 is repeated below for convenience:

IP Boundary Tones	Realization	Meaning
H%	H tone target at end of IP with potential f0 range expansion/raising towards end. Interacts variably with preceding L targets	Particle-final questions; echo questions; disbelief
L%	Strong, separate L target; will form a clear contour on IP-final H tone syllables	Marked statements; negative, corrective, emphatic, obviousness
%	No clear separate pitch target, only realized via phrase-final lengthening, pitch range compression, and optional tonal reduction	Unmarked/general statements; wh-questions; A-not-A questions; particle-final questions; rhetorical questions

Table 11. Inventory of IP-final boundary tones with their realizations and meanings. Same as Table 10 above.

5. Break Indices

Break indices are numeric labels that mark the degree of juncture, which is the perceived amount of 'separation' between syllables that is influenced by prosodic effects such as pitch, duration (i.e., segmental lengthening or shortening), and silence, as well as surface realizations of segments. In the system proposed here, break numbers along with the diacritics explained below can also help clarify non-canonical tonal behavior due to segmental loss or nearby focused elements.

The basic break indices in this system are 0, 1, 2, 3, and 4. Break 0, as the smallest, is used when there is a perception of segmental loss or shortening between two syllables. Break 1 marks the juncture between syllables/morphemes within an Accentual Phrase (AP). Break 2 corresponds to the juncture between APs; Break 3 to the juncture between intermediate phrases (ips); and Break 4 to the juncture following an Intonational Phrase (IP). In all cases presented here, a linguistically-aware native speaker consultant helped confirm the perceived juncture.

There are three diacritics, "m," "p," and "-," which can follow the numerical break index. Xm indicates a break that is somehow mismatched in cues and perception; i.e. the break should be perceived as equivalent to the numerical index used, but lack other qualities associated with such a break. Xp is used to mark speech errors or perceived disfluencies of multiple kinds. Lastly X- is used in cases of labeler uncertainty or when the break between two syllables is perceived as ambiguous between two levels. While all breaks other than 0 could conceivably be combined with each of the three diacritics, only a subset of these potential combinations have clear examples in the data, due in part to the laboratory speech style used by most speakers. In the

following sections, each break index will be explained in turn, and under each index the confirmed variants thereof will be introduced. As of yet, only the following indices with diacritics have found use: 2m, 1p, 2p, 3p, 1-, 2-, and 3-. With future research, especially using more naturalistic speech, the inventory of used index+diacritic combinations will most likely expand.

5.1. Break 0

Break 0 is the smallest juncture, and must be perceived as shorter than the average break between syllables inside an AP, usually due to some segmental reduction or deletion. Such reduction may occur on either the preceding or following syllables of the juncture. Despite this cheating of segments, there is no percept of a speech error. Additionally, because the syllables are shortened in these cases, there may be more extreme pitch delays than expected across a Break 1. In the current proposal, none of the break diacritics (Xm, Xp, X-) co-occur with 0.[18]

In the following examples (Figs. 53 and 54), two speakers produced the same phrase. The utterance in Fig. 53 shows a Break 0 between an adjective and the grammatical linking particle, 的 [ɣəʔ]. Here, the deletion of [ɣ] and the +H plateau contribute to the percept of a Break 0.

[18] 0m and 0- both seem impossible, as there are no real cues to 0 other than segmental deletion and shortening, and there is no smaller break to be ambiguous with.

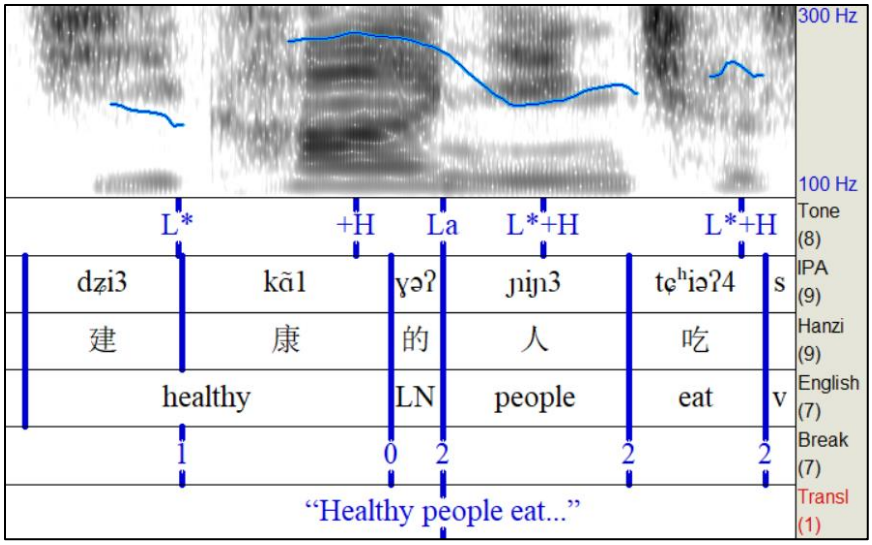

Figure 53. Annotated Shanghainese phrase "Healthy people eat…" showing a Break 0 between the second and third syllables, as evidenced by segmental reduction and plateaued +H on the second and La on third syllables. Same as Fig. 24 above. Speaker: 7c.

In comparison, the same utterance produced by another speaker (see Fig. 54) shows a regular break 1, producing the expected pitch contour for the phrase. While the expected [ɣ] of 的 'ADJ' has been realized as [ŋ] due to the preceding nasal vowel, the speaker nevertheless has a clear onset for the syllable, unlike the speaker above.

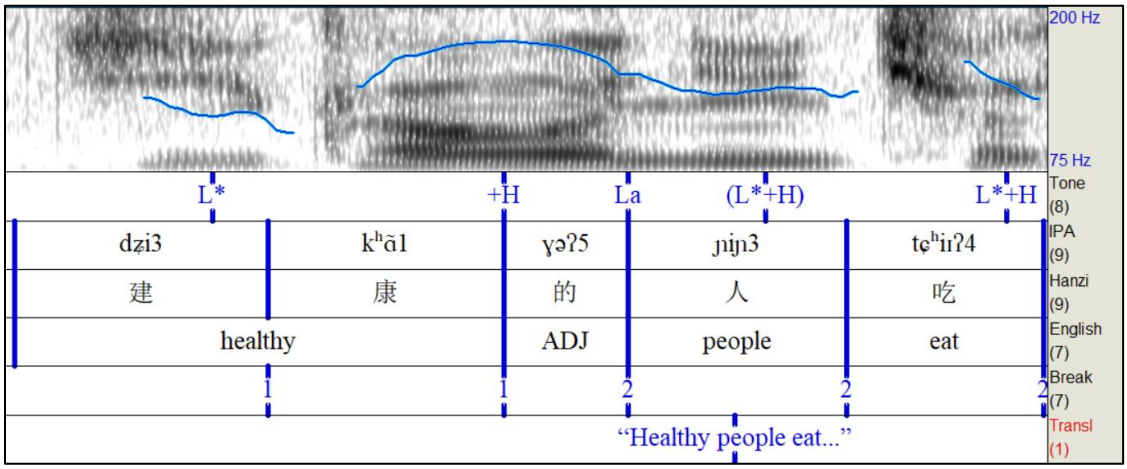

Figure 54. Annotated Shanghainese phrase "Healthy people eat…" showing a normal 1 between the second and third syllables, as evidenced by the full realization of segments, and normal tonal interpolation between the +H on the second syllable and the L* of the fourth. Same as Fig. 25 above. Speaker: 6c.

While Fig. 53 above illustrated segmental deletion following a 0, the next example shows the opposite. The utterance in Fig. 55 shows a 0 juncture between the two words 就 [dʑiɤ] ‘just’ and 要 ‘want’ [iɔ], triggered by the reduction of [ɤ] and the delay of the L* target. Again, the second figure (Fig. 49) in this pair shows Break 1 in the same context for comparison.

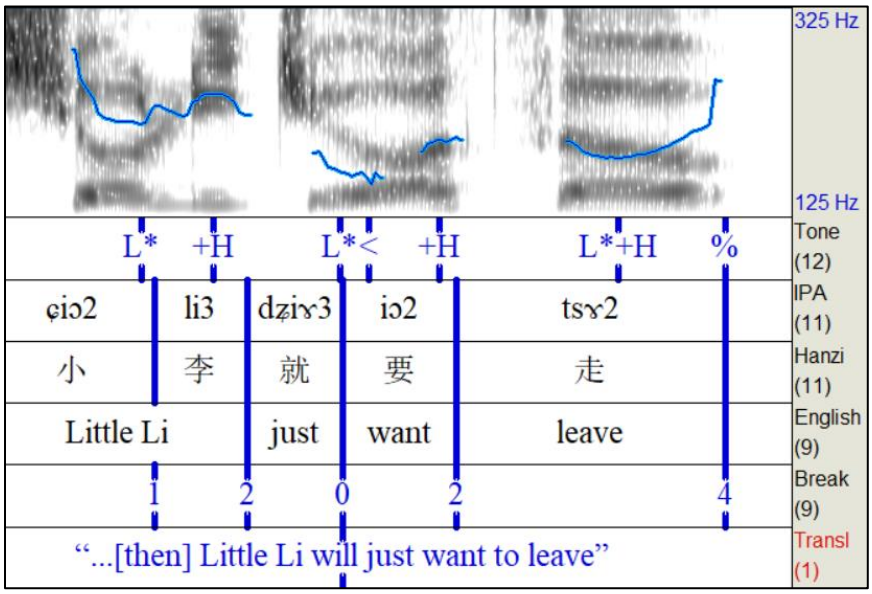

Figure 55. Annotated Shanghainese phrase “…[then] Little Li will just want to leave,” showing Break 0 between the third and fourth syllables, 就 ‘just’ and 要 ‘want.’ Same utterance as in Fig. 32. Speaker: 7c.

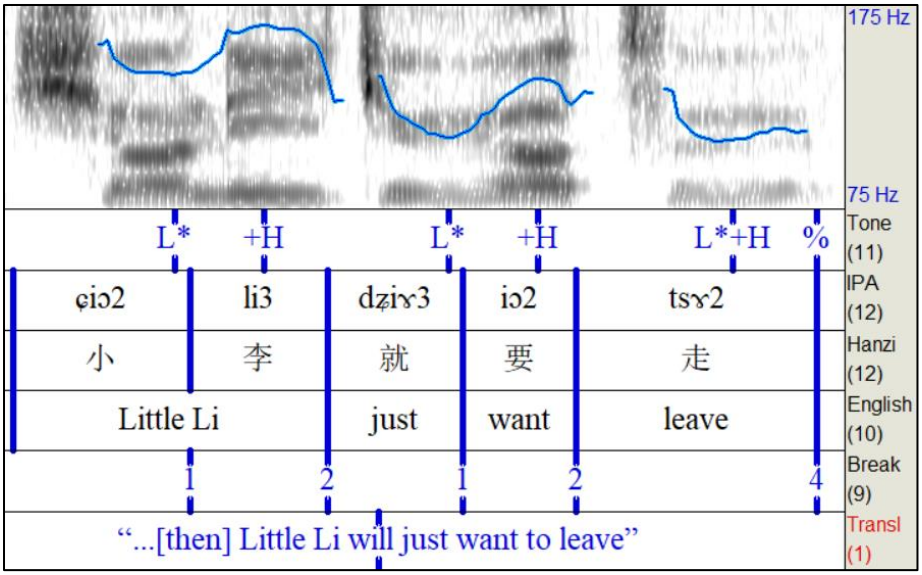

Figure 56. Annotated Shanghainese phrase "…[then] Little Li will just want to leave," showing Break 1 between the third and fourth syllables, 就 'just' and 要 'want.' Speaker: 2c.

5.2. Break 1

Break 1 represents the percept of the juncture between syllables within the same AP. As such, there is no segmental lengthening preceding or following the juncture, and clear predictions are made about how tones should be realized on either side (i.e. spreading or interpolation). As of yet, there are only examples of 1p and 1-.

5.2.1. Break 1p

As a p-type break, 1p is only used to mark instances where there is a percept of disfluency or speech error. This break in particular is used for disfluent junctures that are accompanied by no lengthening, involving an abrupt cut-off in speech and potential segmental loss. In Fig. 57 below, the speaker has two 1p type disfluencies; neither has actual loss of segmental content, but both contain inserted glottal stops, cutting the previous syllables short.

Fig. 58 shows the same sentence said fluently by another speaker, with Break 2 labeled between the relevant syllables.

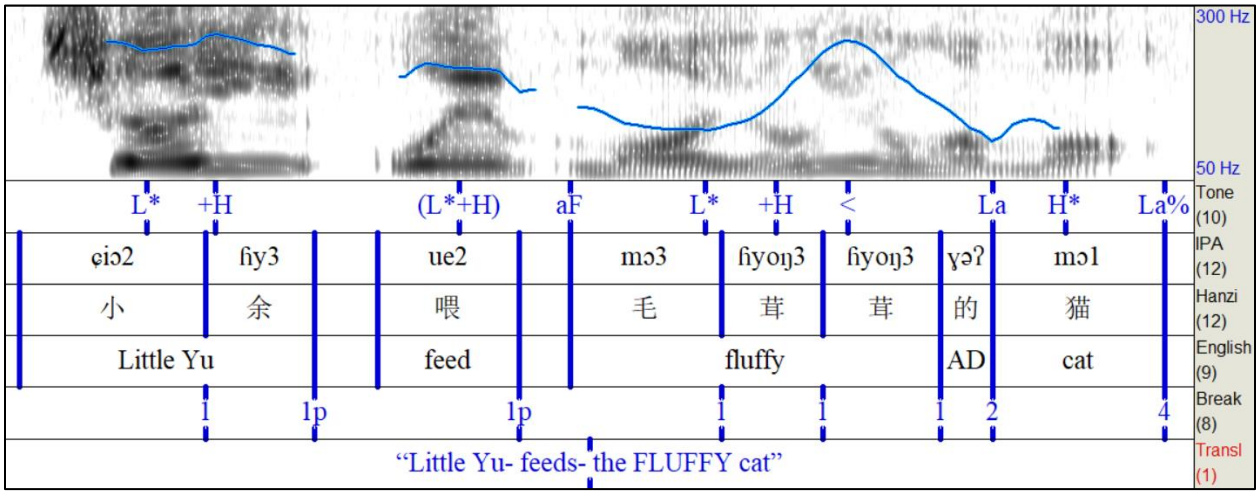

Figure 57. Annotated Shanghainese sentence "Little Yu- feeds- the FLUFFY cat," showing two instances of a 1p, where the speaker abruptly cuts off during a disfluency, inserting glottal stops. In both cases here, the disfluencies are followed by silences. Speaker: 6c.

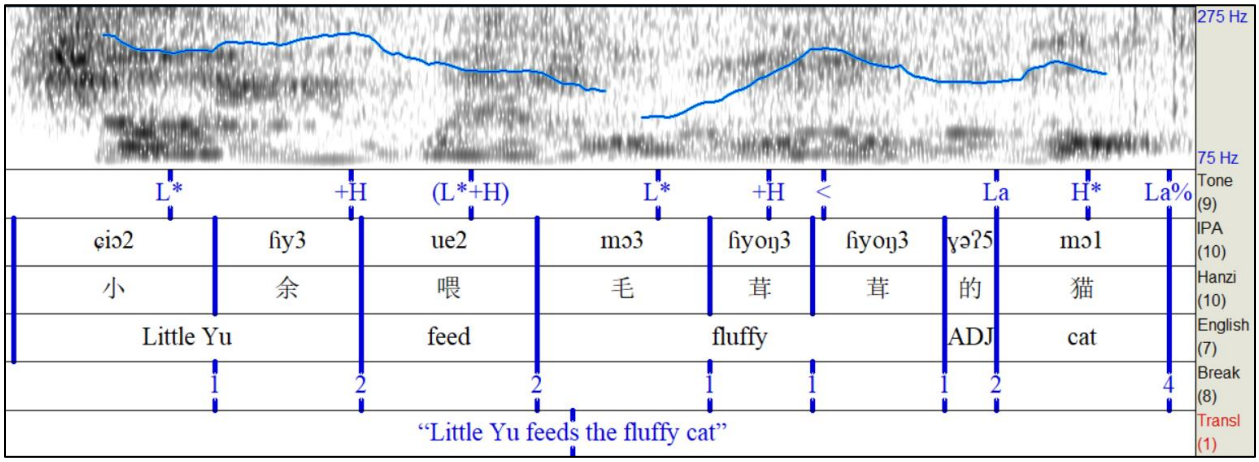

Figure 58. Annotated Shanghainese sentence "Little Yu feeds the fluffy cat," showing no disfluencies. Speaker: 9c.

5.2.2. Break 1-

Break 1- is used to indicate a juncture that is perceived as being between a normal 0 and normal 1 level. In Fig. 59 below, we can compare the AP-internal breaks in the first two APs, 我的 ‘my’ and 妈的 ‘cat’s’

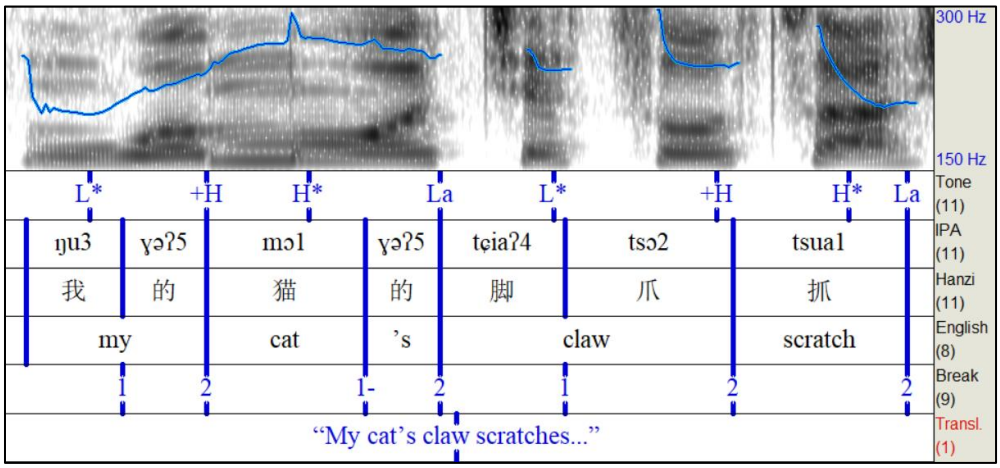

Figure 59. Annotated Shanghainese phrase ‘My cat’s claw scratches…’ showing the difference between Breaks 1 and 1- in the first two APs. Same as Fig. 17 above. Speaker: 7c.

The first AP, with a 1 between its two syllables, has a full realization of the possessive particle’s (的) onset, [ɣ]. In the second AP, the same morpheme is realized without an onset, but without shortening the syllable’s length at all, giving the impression of a shortened juncture.

5.3. Break 2

Break 2 represents the percept of the juncture between syllables across an AP boundary, which can be seen in many figures throughout the dissertation. Break 2 is currently the only break that has been combined with all diacritics: 2m, 2p, and 2-.

5.3.1. Break 2m

2m indicates a juncture that has properties of both an AP-internal boundary and a cross-AP boundary, but instead of an AP-internal percept of the juncture, a cross-AP amount of juncture is felt. In these cases, similar to 1m above, there are still traces of non-spread tonal contours on a syllable-by-syllable level despite overall contours that suggest a more AP-like grouping of multiple syllables. This kind of break generally occurs between two syllables that would have otherwise been monosyllabic APs on their own. Again, this could also be used to describe other gradient spreading cases as described by Zhang & Meng (2016) and Zhang & Yan (2018).

In Fig. 60 below, a 2m occurs between the first two APs, 我 'I,' and 老 'very.' In addition to both being monosyllabic elements, they also precede the focused AP 欢喜的人 'person [that I] like,' which, as was shown in Section 4.5., can lead to tonal reduction in preceding APs. An L and H target are both distinct on each of the two monosyllables; however, there is still a larger trend over the two syllables from the lowest L target on 我 'I,' to the highest H target on 老 'very.' While it may be possible to disregard the larger tonal trend as irrelevant to the phrasing at first glance, the predictions from IP-initial pitch range expansion (IPRE) complicate the analysis. IPRE states that the first AP should have the highest pitch target unless there is focus later in the sentence, or if there is a +H)(H* sequence over the first cross-AP boundary. While focus is present here, it is not on the second AP, which leaves the question as to why it has a higher H target than the preceding, IP-initial AP. Thus, the tonal behavior here suggests that these two syllables are in limbo between being a single, disyllabic AP or being two, monosyllabic APs, resulting in the underlying contours being realized on the surface while an overall L* to +H pattern is apparent over the two syllables.

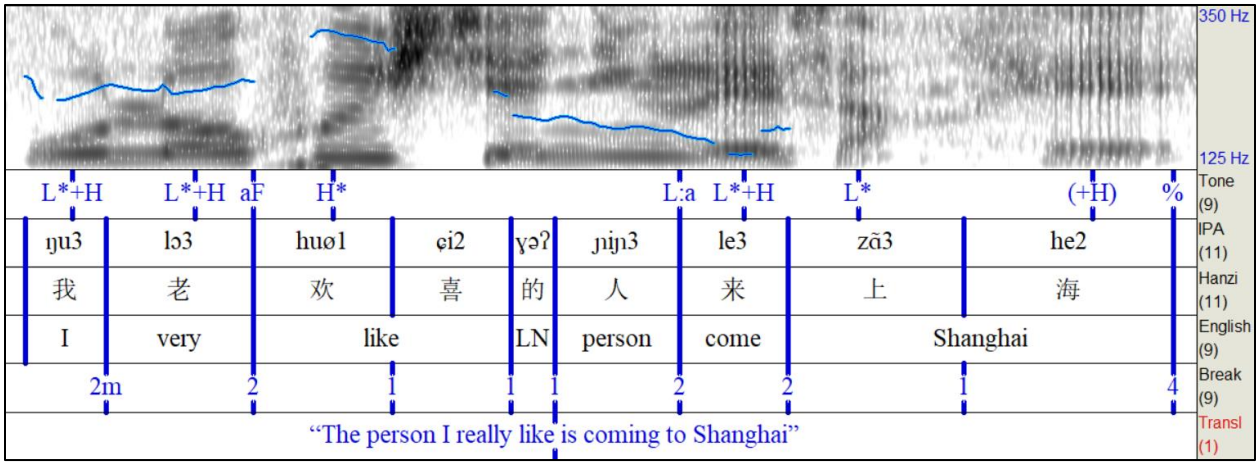

Figure 60. Annotated Shanghainese sentence "The person I really like is coming to Shanghai," showing a 2m between the first two syllables, as evidenced by the individual L*+H contours on each syllable despite the overall low to high pitch track across the two syllables. Speaker: 7c.

In contrast, in Fig. 61, the same sentence shown in Fig. 60 is produced by a different speaker. Here, a normal Break 2 has been labeled between the first syllables, 我 'I' and 老 'very.' Here we can see that the first AP is the highest pitch in the utterance, and the second syllable, though partially tonally reduced, is clearly in its own AP.

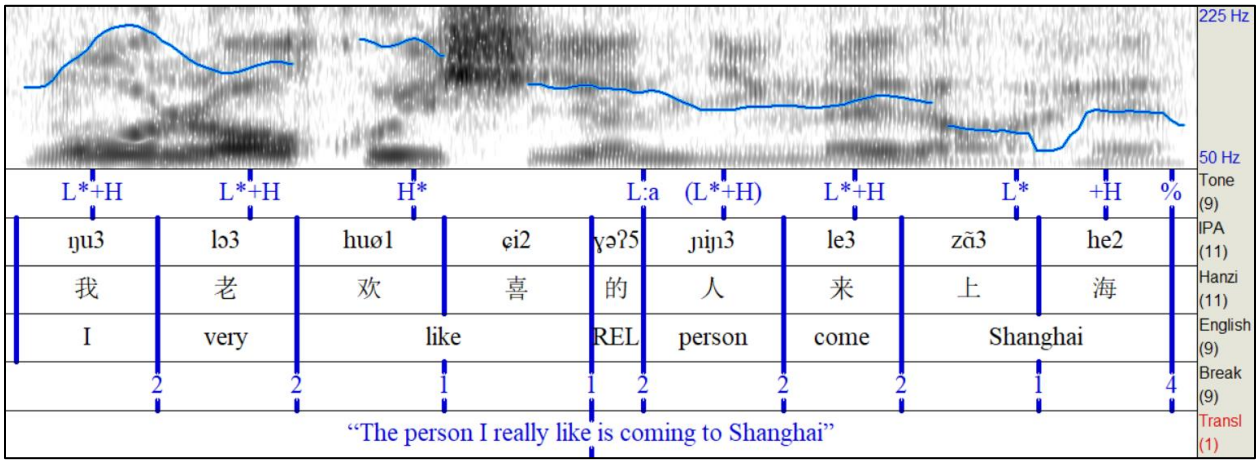

Figure 61. Annotated Shanghainese sentence "The person I really like is coming to Shanghai," showing a 2 between the first two syllables. Same utterance as Fig. 21 above. Speaker: 2c.

5.3.2. Break 2p

As a disfluency break, 2p indicates junctures where there is neither an abrupt cut-off nor segmental lengthening before a speech error. As is evidenced in Fig. 62, 2p may or may not be followed by a silence. Over the first 2p in the figure, between 猫的 ‘cat’s’ and 脚爪 ‘claw,’ there is an intervening silence and pitch reset (as L* appears higher than the previous +L), while on the second, between 脚爪 ‘claw’ and 抓 ‘scratch,’ no silence is present beyond the closure of the stop in [ts]. This can be compared to the following pitch track in Fig. 63, showing a normal, fluent production of the same sentence.

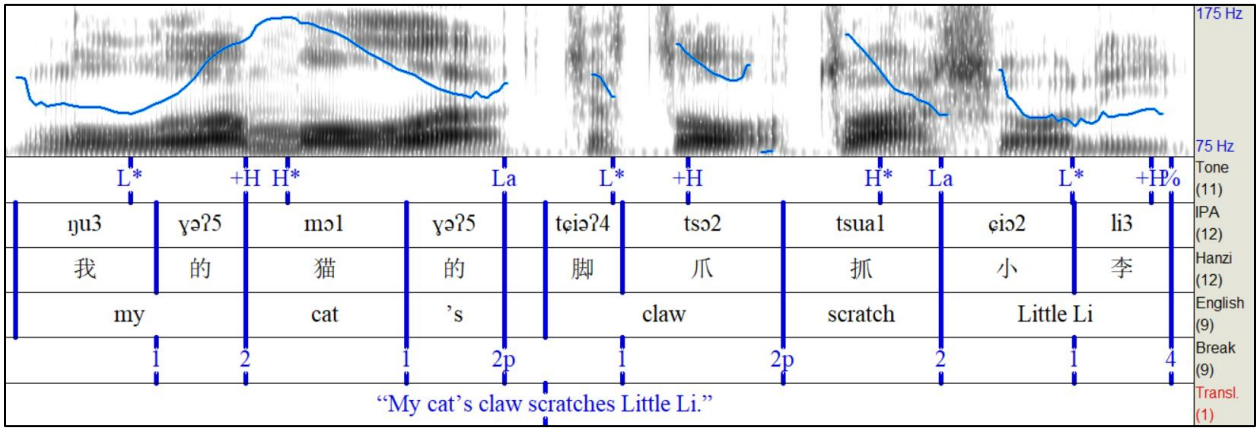

Figure 62. Annotated Shanghainese sentence “My cat’s- claw- scratches Little Li,” showing two 2p disfluencies. In both cases, no segmental material is lost, and no syllables are lengthened. In the first case, between 猫的 ‘cat’s’ and 脚爪 ‘claw,’ a silence follows the disfluency, but between 脚爪 ‘claw’ and 抓 ‘scratch,’ no silence is present despite a clear percept of disfluency. Speaker: 1c.

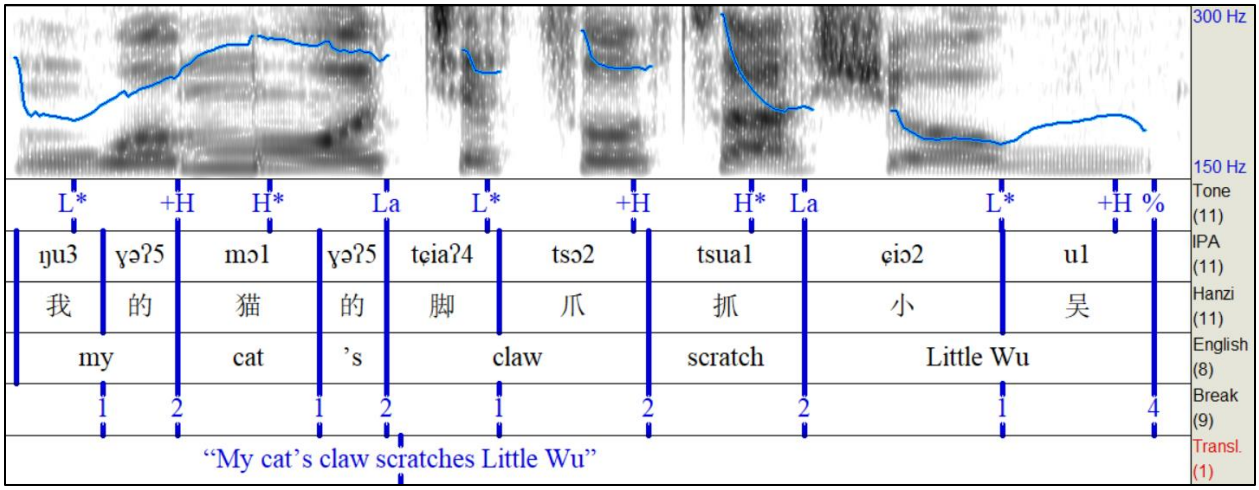

Figure 63. Annotated Shanghainese sentence "My cat's claw scratches Little Wu," showing fluent production. Same utterance as in Figs. 17 & 57. Speaker: 7c.

5.3.3. Break 2-

Break 2- is used to show a juncture perceived as being between a normal 1 and a normal 2 level. In the Fig. 64 below, a 2- was labeled between the second AP 骂我 'scold I' and the third AP, 老 'very.' The pronoun 我 'I' has been shortened considerably, but there are clear tonal indications as to a AP boundary with the immediately following L*+H on 老 'very.' In this case, the tonal cues win out perceptually, making the juncture feel like a shortened Break 2. This can be compared with the following Fig. 65, where a normal 2 has been used in labeling the same boundary.

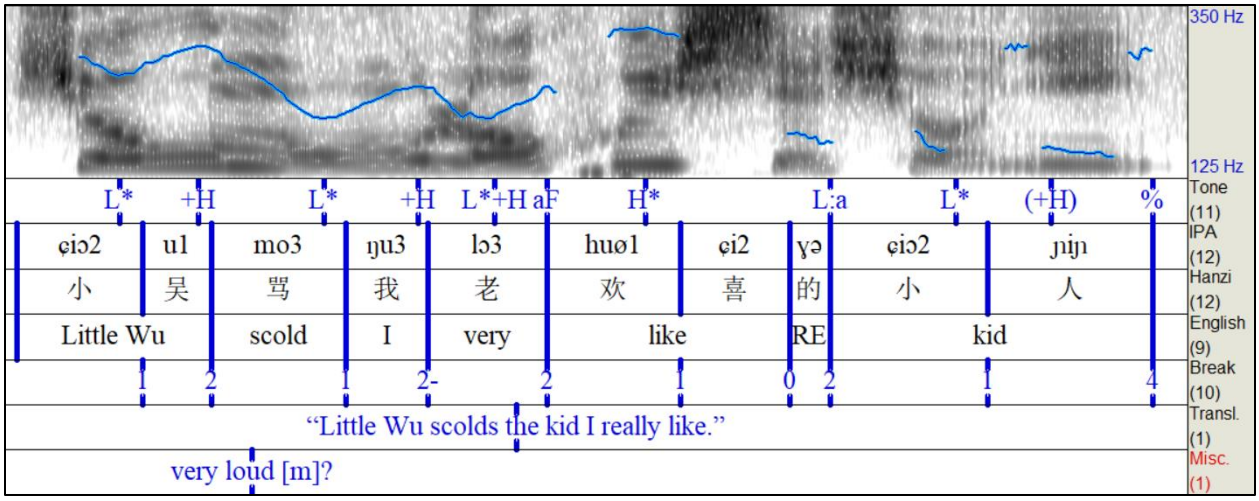

Figure 64. Annotated Shanghainese sentence, 'Little Wu scolds the kid I really like,' showing a 2- between the second AP, 骂我 'scold I,'[19] and the third, 老 'very.' The segmental loss with clear tonal pattern gives the impression of a shortened cross-AP boundary. Same utterance as Fig. 19 above. Speaker: 7c.

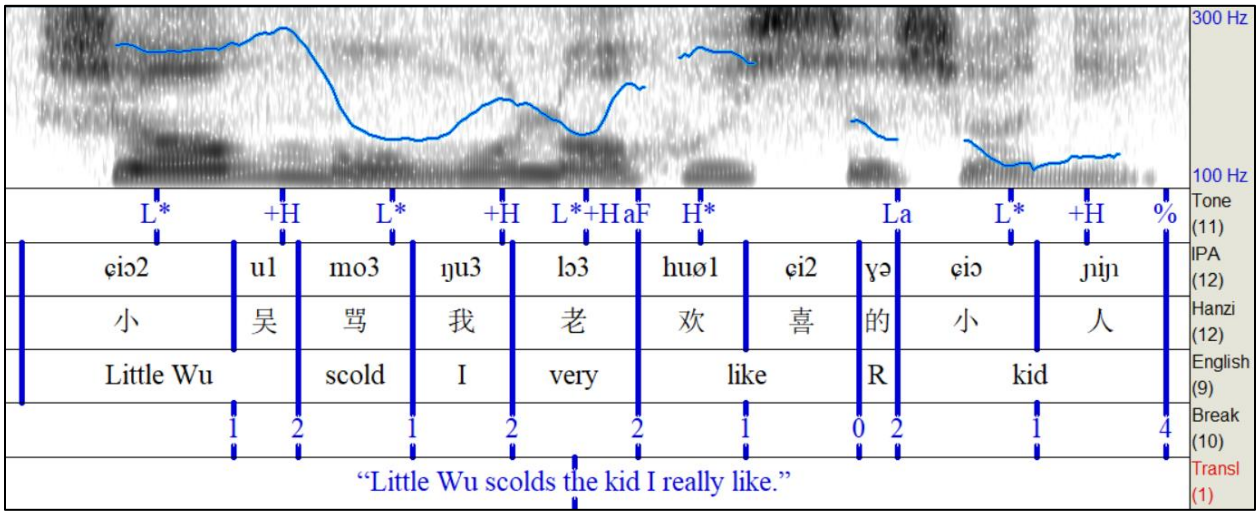

Figure 65. Annotated Shanghainese sentence 'Little Wu scolds the kid I really like,' showing the use of a standard 2 between the second AP, 骂我 'scold I,' and the third 老 'very.' Same utterance as Fig. 18 above. Speaker: 6c.

[19] Note that this AP spans a relative clause/CP boundary, with the matrix verb and embedded subject forming a single unit. This was by far the most common phrasing for this sentence, instead of the perhaps more expected phrasing of two separate APs for 骂 'scold' and 我 'I.'

5.4. *Break 3*

Moving up in size, Break 3 represents the percept of the juncture between syllables across an ip boundary. In this dissertation, 3p and 3- have both found uses, as explained below.

5.4.1. *Break 3p*

As the largest disfluency break, 3p is used when a disfluency is accompanied by significant lengthening of some segmental material, regardless of what side of the juncture the segments are on. In Fig. 66 below, we see two disfluencies: the first is a 2p between 去 'go' and 看 'see,' while the second, a 3p, occurs between 看 'see' and 苏州 'Suzhou.' While the syllable preceding the 2p has not been lengthened, and even undergoes tonal reduction, the syllable preceding the 3p is significantly lengthened in comparison with its surroundings, and has a full tonal contour. A small silence follows the 3p disfluency here, though this is not necessary for diagnosis of 3p.

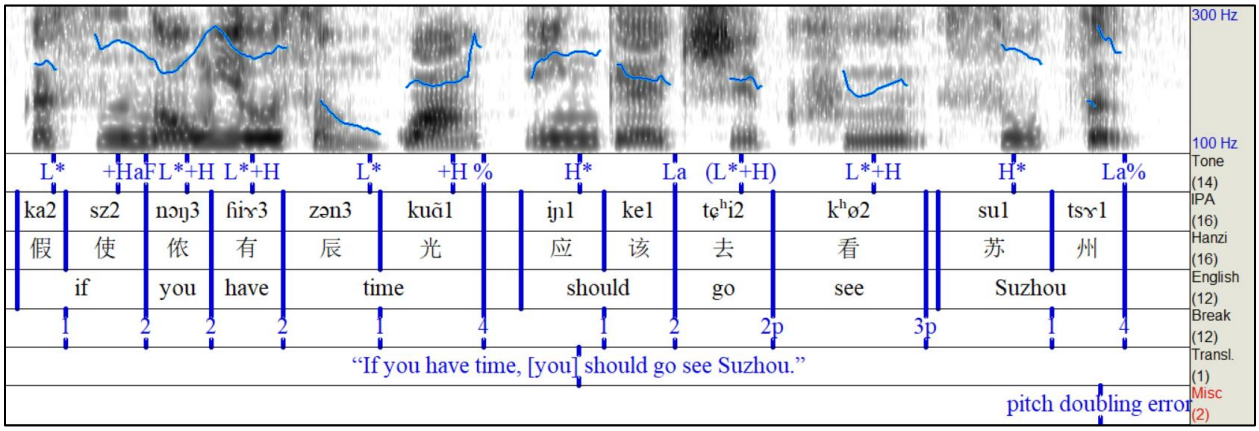

Figure 66. Annotated Shanghainese sentence "If you have time, [you] should go see Suzhou," showing two disfluencies. Between 去 'go' and 看 'see,' a 2p is used to reflect a percept of disfluency with no significant lengthening or segmental reduction. Between 看 'see' and 苏州 'Suzhou' a 3p is labeled to reflect significant segmental lengthening, allowing for a full L*+H contour on 看 'see.' In this case, the 2p is not followed by silence, while the 3p is. Speaker: 3c.

In contrast, Fig. 67 shows the same sentence as in Fig. 66, produced by a different speaker with a normal 2 break between 看 ‘see’ and 苏州 ‘Suzhou.’

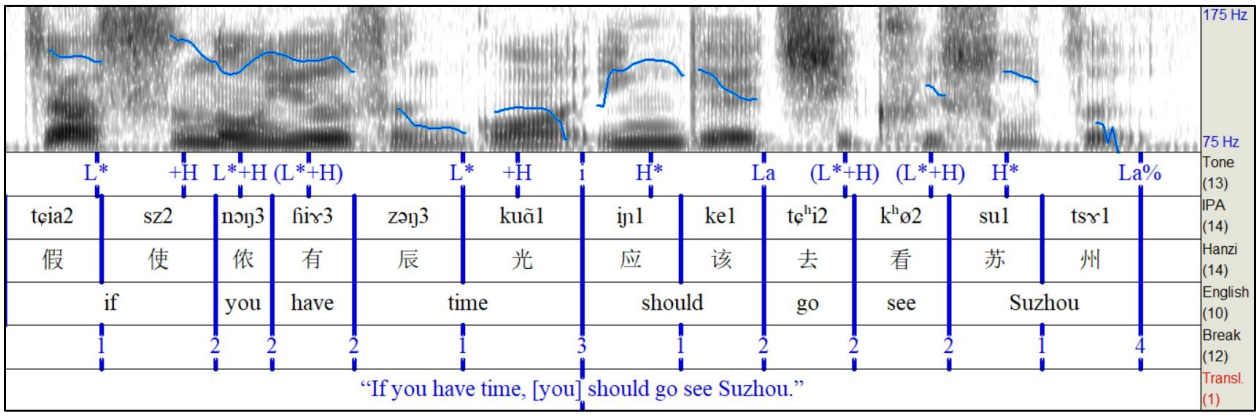

Figure 67. Annotated Shanghainese sentence “If you have time, [you] should go see Suzhou,” showing fluent production, with an ip-level juncture 3 between the clauses. Note that both monosyllabic verbs in the second clause, 去 ‘go’ and 看 ‘see,’ have undergone tonal reduction. Speaker: 1c.

5.4.2. Break 3-

If a fluent juncture is perceived as being between an ip boundary and an AP boundary, the label 3- is used. The three following figures (Figs. 57-60) compare Break 2, Break 3-, and Break 3, explaining how they can be distinguished based on perception, and what factors potentially contribute to that perception.

In Fig. 68, a Break 2 is labeled between the two clauses, suggesting that the degree of juncture here corresponds to the typical juncture for an AP boundary. Even though this break could have an ip break (as in Fig. 59), we can see no lengthening on the clause final syllable 来 ‘come,’ and the L* on the first syllable of 阿拉 ‘we’ is a local minimum, as predicted for an ip-internal AP. These factors show that both clauses have been phrased into a single ip.

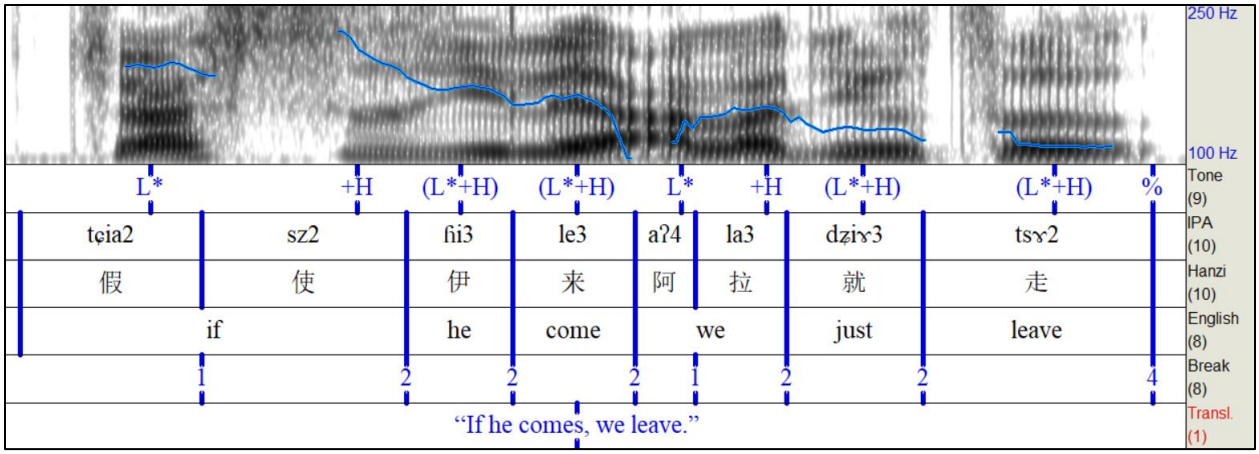

Figure 68. Annotated Shanghainese sentence 'If he comes, we leave,' uttered at a fast speech rate, showing a 2 between the two clauses, as evidenced by the lack of pitch reset, and lack of lengthening on the clause-final syllable 来 'come.' Speaker: 1c.

Fig. 69 shows an example of 3-, the shortened break, between the same two clauses shown in Fig. 68. Here, pitch reset is visible across the juncture between 来 'come' and 阿拉 'we,' where L* is higher than the preceding reduced tone. While this pitch reset results in the percept of a larger juncture than a Break 2, the syllable 来 'come' is not lengthened to the degree expected of an ip-final AP. Because of this, the perceived juncture is between a Break 2 and a 3, meaning 3- is the most appropriate break index.

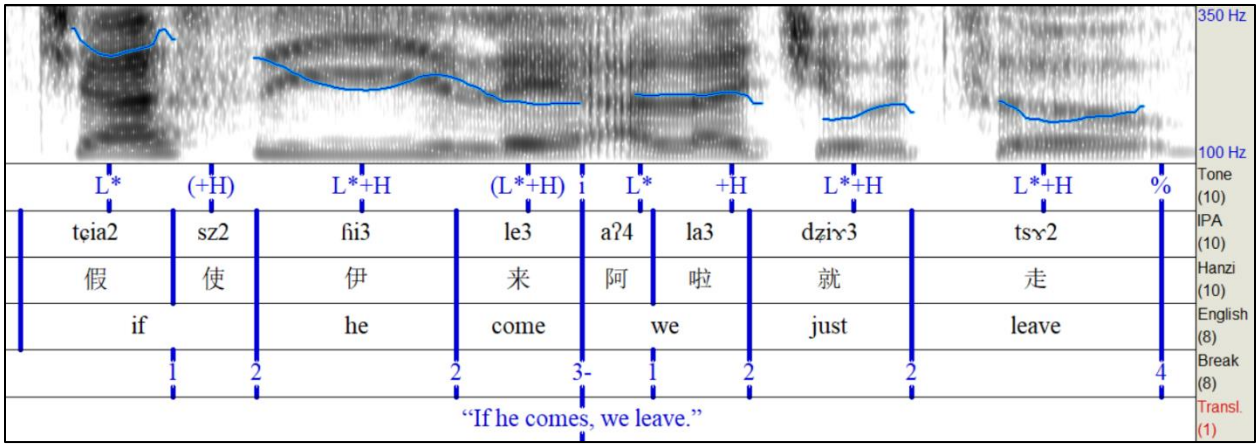

Figure 69. Annotated Shanghainese sentence 'If he comes, we leave,' showing a 3- between the two clauses, indicating a shortened ip juncture. The weaker degree of lengthening on the clause-final syllable 来 'come' combined with the following pitch reset give the impression of a shortened ip break. Speaker: 7c.

In the last of these comparisons, Fig. 70, we see the same sentence, but with a normal 3.

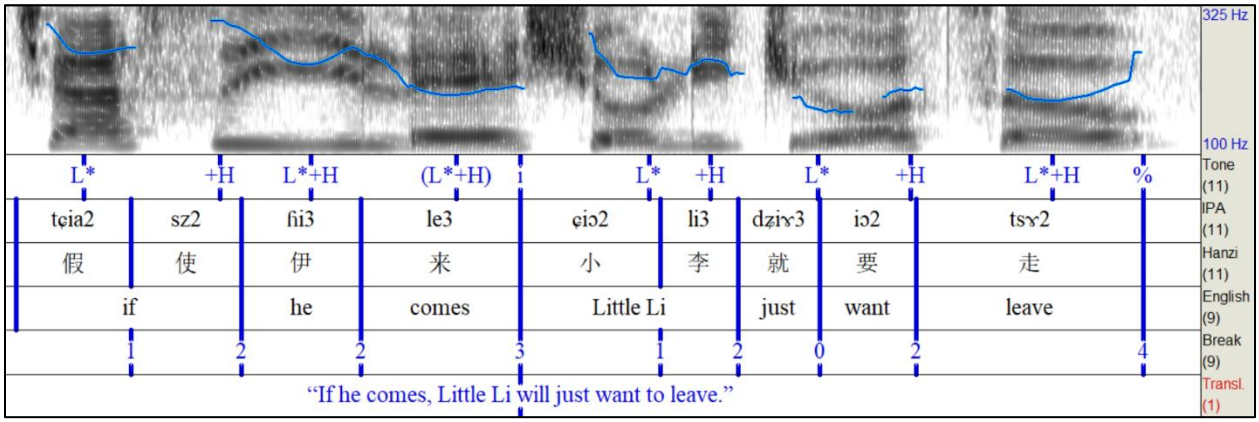

Figure 70. Annotated Shanghainese sentence 'If he comes, Little Li will just want to leave,' showing a 3 between the two clauses. Same as Fig. 32 above. Speaker: 7c.

5.5. Break 4

Break 4 is the largest break, corresponding to the juncture at the end of an Intonational Phrase (IP). As such, it involves significant segmental lengthening, tonal compression of the preceding syllable, and crucially, must it be followed by a silence—note that this is the only non-

disfluency break that can co-occur with a silence. Examples of Break 4 are present in all full-sentence data in previous figures. In Fig. 35, an IP-break has been used sentence-medially between two clauses. While the Break 4- could potentially be needed in later iterations of this model, as of the time of writing, no datum seems to need such a label.

5.6. Summary of Break Indices

Over this section, the break index system has been introduced, with its five level of juncture and three diacritics. The following table summarizes the various combinations presented and discussed as part of the system.

Break	**Xm** *Mismatch*	**Xp** *Disfluency*	**X-** *Shortened*
0 AP-internal break with segmental loss	N/A	N/A	N/A
1 AP-internal break	*1m* *Tonal cues show a clear AP boundary, but the degree of juncture is like an AP-medial word boundary*	1p Shortened disfluency with no lengthening and potential segmental loss.	1- Juncture perceived as between 0 and 1
2 Cross-AP break	2m Percept of cross-AP break; overall trend of tone spreading but with strong traces of underlying contours on each syllable	2p Disfluency with no discernable shortening or lengthening	2- Juncture perceived as between 1 and 2.
3 ip boundary break	N/A	3p Disfluency with significant lengthening	3- Juncture perceived as between 2 and 3.
4 IP boundary break	N/A	N/A	*4-* *Juncture perceived as between 3 and 4*

Table 12. Summary of break indices introduced in this dissertation. N/A indicates break index+diacritic combinations that are not compatible for Shanghainese. Italicization indicates potential break indices that were not exemplified in this dissertation, but could be useful in further research on Shanghainese junctures.

6. Concluding Remarks

6.1. Summary of Findings

The model presented above proposes a pitch accent analysis of Shanghainese with three levels of phrasing above the syllable/morpheme: the accentual phrase (AP), the intermediate phrase (ip), and the intonational phrase (IP). The AP, the lowest level of phrasing, generally consists of a word and any following non-lexical elements—most commonly two to three syllables, but ranging from one to five in the data here. The AP is the domain of the three lexically contrastive pitch accents and their associated phrase-final boundary tones, H* La/L:a, L*+H (La/L:a), and L* LHa, where the initial syllable of the AP is prominent, and thus always is the lexical origin and host of the pitch accent. In APs with simple pitch accents, H* and L*, the boundary tones (La/L:a or LHa, respectively) are always realized, but with the bi-tonal pitch accent, L*+H, the boundary tone La/L:a is only realized in APs longer than two syllables.

The two AP-final boundary tones are La/L:a and LHa. La is realized after H* and L*+H as a single low target aligned to the end of the AP. It varies (rather freely) with L:a, which spreads from the end of the AP leftward until a syllable with an assigned tone is met. LHa is realized as a sharp rise over the last syllable of an AP, which is always preceded by a low plateau from the L* that it co-occurs with.

Across AP boundaries, within the same ip, the prominence imbalance between strong AP-initial tones (H* and L*) and weak non-initial (+H, La/L:a) tones surfaces as the extreme realization of the prominent tonal targets; i.e. whenever two identical tone targets are adjacent across an AP boundary the second, AP-initial tone is either higher than a previous H target (in

the case of H*) or lower than a previous L target (in the case of L*). This has the effect of lending salience to AP edges, as every AP-initial syllable is either a local pitch maximum or minimum.

Above the AP, the intermediate phrase (ip) consists of at least one AP, but generally contains several. There are no tones associated with the ip, but it can still be readily identified by several prosodic effects: 1) the percept of a larger juncture with no actual silence in at the boundary location; 2) its lengthening effect on ip-final syllables; and 3) pitch reset seen on the next AP. Optionally, tonal reduction may also apply to final syllables in an ip, despite segmental lengthening.

The largest level of phrasing, the Intonational Phrase (IP), contains minimally one ip, and contains up to two in the data considered here. IP-initial pitch range expansion (IPRE) marks the beginning of IPs, generally affecting the first AP and ensuring that its high target (whether H*, +H, or LHa) is the highest in the IP. IPRE can sometimes override the prominence relationship between +H and H*, causing an H+ to be higher than a following H*. This kind of extreme IPRE was encountered only in disyllabic IP-initial APs in this data.

In addition to IPRE, the IP is also obligatorily marked with one of three final boundary tones: H%, L%, or %. H%, the high IP-final boundary tone, was found only in information-seeking, particle-final yes/no questions. In these instances, the H% could interact with low tones in multiple ways, dependent on speaker and context. All speakers allowed for the ‘combination’ of La and H%, which is realized as level mid tone. A minority of speakers also allowed for deletion of preceding La targets, with pitch gradually rising from the nearest high target to the end of the sentence. Lastly, in IP-final APs with four syllables, all speakers displaced the La

target back on to the third syllable so that H% could be realized as a gradual rise over the last syllable.

L% is the low IP-final boundary tone, which is used in instances of correction, negation, emphasis, or otherwise emotionally marked contexts. Because of the few contexts in the data that elicited such a boundary tone, it was only identified after +H and LHa tones, and so its behavior with the low tones +L and La can only be guessed at this point. When appearing after a high target on a disyllabic AP, the tones compress onto the last syllable, and L% causes a sharp falling contour. In trisyllabic contexts, the H target is moved earlier to accommodate the L%, but the final contour still has a definitive fall.

As a toneless boundary event, % does not affect preceding lexical or phrasal tones, though sometimes IP-final syllables have markedly reduced tone, just as in ips. Additionally, % also causes lengthening of the IP-final AP and will entail pitch reset if speech continues. Thus, the only reliable difference between the % IP boundary tone and the ip boundary is the existence of silence which always follows the IP-final %. This boundary is used in most general, unmarked statements, but is also seen in questions of all types investigated (i.e. particle-final yes/no, A-not-A yes/no, wh-elements, rhetorical).

The following two figures summarize the model. Fig. 71, repeated from Fig. 1 above, shows the proposed intonational hierarchy of Shanghainese, while Table 13 shows a list of all tone tier annotations for Sh_ToBI.

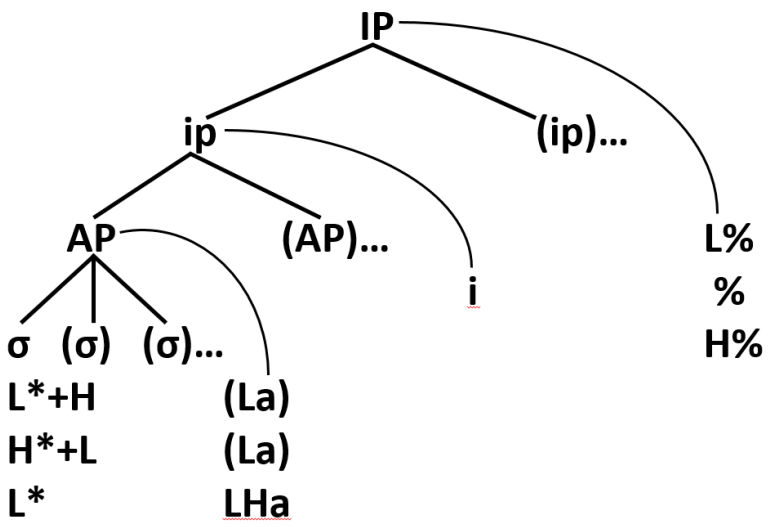

Figure 71. Schematic representation of Shanghainese phrasal hierarchy and intonation; same as Fig. 1.

Tone Tier Annotations	Use	Alignment
L*+H, H*, L*	Lexical pitch accents	Each tone aligned to pitch minimum/maximum on phonological host
La, L:a, LHa	AP-final boundary tones	Aligned to end of AP
i	ip-final boundary marker	Aligned to end of ip
L%, %, H%	IP-final boundary tones[20]	Aligned to end of IP
aF	Narrow focus	Aligned to beginning of narrowly focused AP
aC	Corrective focus	Aligned to beginning of correctively focused AP
<	Delayed pitch target	Aligned to an f0 maximum/minimum that is realized after its phonological host syllable
>	Early pitch target	Aligned to an f0 maximum/minimum that is realized before its phonological host syllable
(T)	Reduced/unclear tone	Parenthesis go around tone on syllable that has been affected by tonal reduction

Table 13. Summary of all tone tier annotations in Sh_ToBI.

[20] See Table 10 above for descriptions of realization and meanings of each IP-final boundary tone.

In addition to the analysis of Shanghainese intonation given above, this dissertation also presents a system of break indices that can be used to categorize the amount of juncture felt between two syllables, as well as encode information about gradient phrasal boundaries and disfluencies. The break system is made up of five levels—0, 1, 2, 3, and 4—which correspond to relative percepts of juncture. Break 0 is used for junctures that seem smaller than a normal AP-internal juncture, often with segmental reduction and/or delayed pitch targets. Break 1 is used for junctures equivalent to AP-internal junctures; Break 2 is used for junctures equivalent to cross-AP junctures; Break 3 is used for junctures equivalent to ip junctures, and break 4 is used for junctures equivalent to IP junctures.

However, there are diacritics, Xm, Xp, and X-, which can be used to explain seemingly mismatched junctures (Xm), speech errors and disfluencies (Xp), or unclear/intermediate junctures (X-). The system here made use of 2m, 1p, 2p, 3p, 1-, 2-, and 3- combinations, with the possibility of more as further more naturalistic data is analyzed.

6.2. Contributions to Existing Literature

First and foremost, this dissertation adds to our typological knowledge of intonational systems, by presenting a novel analysis of a language known for its unique tonal properties. Unlike its more closely related and clearly tonal cousins, Mandarin and Cantonese, Shanghainese seems to have a more pitch-accent like system, with prominence distinctions in syllables, tonally unspecified syllables, and phrasal boundary tones; additionally, multisyllabic words seem to be stored whole by speakers. At the same time, these pitch accents interact with the phonation distinction in consonants, producing five different phonetic contours that speakers increasingly

recognize as distinct, and the process of tone spreading seems productive despite whole-word storage. The combination of these seemingly contradictory facts is most evident in the realization of L*. On one hand, in monosyllabic and disyllabic contexts, speakers are able to categorize the tone of the prominent syllable as one of the rising citation tones, dividing what is treated in a pitch-accent analysis as one category, L*, into four: [34], [23], [44], [12]. On the other hand, L* seems to behave the same in relation to the AP-final phrasal accent La regardless of context—L* is always a local minimum, even in the case of the highest citation tone, T4; thus it appears that all of these L*s do share some properties or features. Additionally, as shown by Tian & Kuang (2019) and Yan et al. (2020), speakers store and access whole word forms that only have one tonal specification.

To show the mixed nature of Shanghainese tone, the parameters used in Jun (2005) for cross-linguistic comparison of prosodic systems are given for multiple languages in Table 14 below. The parameters listed here can be divided into two major types—the first relates to the status of prominence in a language: Is it lexically determined? If so what is the type of specification? Are there post-lexical ways of marking prominence? If so, do these mark heads or edges? The second group of parameters have to do with rhythmic or phrasal units that the language makes use of, both sub-word or lexical (mora, syllable, or foot) and supra-word or post-lexical (AP, ip, IP). As we can see, Shanghainese shares properties of both tone (Cantonese and Mandarin) and pitch accent languages (Swedish and Japanese).

Language			Eng.	Sw.	Jp.	Fr.	Kor.	Cant.	Mand.	SH
Overall Prosodic Type			Stress Accent	Stress-PA	Lex. PA	Stress-Edge	Edge Only	Tone Only	Tone-Stress	Tone-PA
Word-Level Prominence	Lexical	Tone						+	+	+
		Stress	+	+					+	
		PA		+	+					+
	Post-Lexical	Head	+	+	+	+				+
		Edge			+	+	+			+
Rhythmic/ Prosodic Unit	Lexical	Mora			+					
		Syll.				+	+	+	+	+
		Foot	+	+						+
	Post-Lexical	AP			+	+	+			+
		ip	+				+		+	+
		IP	+		+	+	+	+	+	+

Table 14. Cross-linguistic summary of prominence and rhythmic parameters as proposed by Jun (2005). Abbreviations and sources: Eng. = English (Beckman et al. 2005), Sw. = Swedish (Bruce 1982), Jp. = Japanese (Venditti 2005), Fr. = French (Jun & Fougeron 2002), Kor. = Korean (Jun 2005), Cant. = Cantonese (Wong et al. 2005), Mand. = Mandarin (Peng et al. 2005), SH = Shanghainese (current proposal); PA = pitch accent

As already stated, Shanghainese lexical pitch distinctions seem to share properties of both tones and pitch accents, but other than the lexically-specified pitch on each monosyllable, and the higher levels of phrasing in Mandarin (i.e. ip, IP), the system of Shanghainese seems to align more with pitch accent languages, like Japanese and Swedish. This mainly comes from post-lexical prominence, where Shanghainese shows both head marking (primacy of AP-initial syllable's tonal realization) and edge marking (in the case of La/LHa), while Mandarin and

Cantonese do not. Additionally, Shanghainese is the only questionably tonal language presented here to make use of the foot as a prosodic unit—though most normal sentences maybe phrased according simply to syntax, the AP in Shanghainese is still generally only 2-3 syllables. Additionally, longer complex compounds or loan words have been shown to behave according to syllable-based trochaic feet (Duanmu 1995, 1997). Again, this is another way in which Shanghainese differs from strongly tonal languages like Cantonese.

We can also approach Shanghainese from the view of Hyman (2006, 2007), who does not view pitch accent as a valid typological category. Instead, he has two major deciding questions that broadly apply to tone and stress. 1) Does pitch play a role in lexical distinctions? If so, the language is tonal; 2) Does the language require each word to have at least one prominence (obligatory prominence)? If so, the language is stress-accented. Further, one must characterize the relationship between the tone and stress systems if both are present. Beyond this, he provides a few other questions that can help further divide languages, though he makes no claims about how they do or do not correlate with a language having stress, tone, both, or neither.

Language	English	Swedish	Japanese	French	Korean	Cant.	Mand.	SH
Overall Prosodic Type	Stress-accent	Word tone dependent on Stress	Restricted word tone	Postlexical Stress-accent	Post-lexical only	Syl. tone	Syl. tone	Phrasal tone dependent on Stress
Lexical use of pitch		+	+			+	+	+
Obligatory prominence	+	+		+				+
Culminative prominence	+	+	+	+			+	+
Prominence subordination	+	+						
Predictable prominence				+	+[21]			+
Rhythmicity		+	+	+	+			+

Table 15. Cross-linguistic summary of some tone and stress-accent parameters as proposed by Hyman (2006). Sources: English (Beckman et al. 2005), Swedish (Bruce 1982), Japanese (Venditti 2005), French (Jun & Fougeron 2002), Korean (Jun 2000), Cant. = Cantonese (Wong et al. 2005), Mand. = Mandarin (Peng et al. 2005), SH = Shanghainese (current proposal).

Under Hyman's view, Shanghainese would thus be categorized as 'phrasal tone, dependent on stress.' However, it is important to remember that, while I have characterized the word-level intonational behavior of Shanghainese in this typological discussion, Shanghainese speakers still have knowledge of what would normally be considered two different levels of representation in a generative view. If speakers have access both syllable-by-syllable and word-by-word representations of many if not the vast majority of lexical items, it becomes unclear how

[21] Hyman (2006) considers post-lexical edge-marking tones to be 'predictable prominences.'

best to typologize the language. Even though Hyman's typology claims to clarify the ambiguous term 'pitch accent,' Shanghainese falls between classifications in both views—either lexical tone vs. pitch accent on one hand, or syllable tone vs. word tone on the other.

In terms of the literature pertaining specifically to Shanghainese, my master's thesis and this dissertation also present some of the first work looking at sentences as a phonetic whole. In previous research, only dialectological (e.g. Xu et al. 1986) and some generative work (Selkirk & Shen 1990) considered full sentences in describing and analyzing tonal processes—both are fully impressionistic approaches. Conversely, most phonetic work has focused on the production and perception of only one APs/sandhi domain, either in isolation or in carrier sentences. My work splits the difference, in considering phonetic data from a sentence-level, phonological perspective. This allows for the observation and analysis of variation overlooked in earlier phonological studies, while still seeking categorical prosodic events that are generally not the focus of most phonetic research. Because of this intonational approach, my work is the first to formally discuss several phenomena that are not investigated in previous literature about Shanghainese, such as boundary tones, phonetic diagnostics for the intermediate phrase, cross-boundary prominence effects, and tonal crowding strategies.

Certain findings from previous literature were supported, though others were not found to be tenable. For instance, as the speakers analyzed here were speakers of Middle Shanghainese, the tonal patterns found by Chen (2008) that suggest a final low target at the end of a sandhi domain/AP were confirmed and analyzed as La; additionally, her characterization of this tonal target as 'weak' is also spot on, as shown by its behavior when preceding L* and H%. Takahashi (2013) and Zhang & Meng (2016) both found that right-dominant sandhi was not fully

neutralizing, and instead characterized the process as tonal reduction—this was also the case with my speakers, where the realization of reduced tones varied widely dependent on the both syllable's citation tone and intonational context, indicating the non-neutralizing behavior of tonal reduction. In regards to the 'tonal displacement' pattern exhibited by T5/L*LHa, previous work had given a range of factors governing the application of displacement versus the normal spreading, such as tone of the second syllable in the AP, morphological structure, and lexical specification. In my data, only the last point held: while some APs that were T5-initial were said by some speakers with 'tonal extension' (L*+H) and said by others with 'tonal displacement' (L*LHa), no speaker varied within themselves on the same lexical item; thus it my speakers have lexically determined whether a given voiced-initial, glottal-rhymed syllable has a L*+H or L*LHa pattern. This supports Zhang & Meng's (2016) finding that T5 tonal displacement is not as productive as the extension seen in other tones.

Selkirk & Shen (1990) were found to be generally correct about the determination of sandhi domains/APs—the left edge of lexical words corresponds to the left edge of a domain boundary. However, as an idealized phonological study, it is deterministic, and does not allow for the variability in phrasing seen both across and within speakers. Beyond this, as shown in my master's thesis and here, their definition of the 'major phrase' is untenable. While their major phrase algorithm helped to explain how non-lexical syllables can initiate an accentual phrase (i.e. via initiating a larger, lexical XP), it did not correlate with any focal effects (as they predicted) or perception of juncture (as I would expect). Additionally, there was no occurrence of their reported post-focal effects in any of my data.

6.3. Limitations of Dissertation

Considering first the data itself, there are weaknesses in the range of data collected, making analyses of certain phenomena difficult. This is especially true in the case of the IP-final boundary tones H% and L%. Additionally, more natural free-flowing speech (e.g. media broadcasts, storytelling, conversation) has not been annotated with the system described here, which is generally a part of developing full ToBI systems. Thus, it is possible that certain phenomena described here are a byproduct of laboratory-style speech, or conversely, that certain phenomena only occur in other discourse styles and thus were not noted in the dissertation. Controlled focus data is also lacking; while some speakers were recorded with focus elicited via a dialogue or other clear context, there are many instances where focus was a result of disfluency or the procedure of reading minimally different sentences, even when randomized. Thus, while it is clear in my data that corrective and narrow focus have different realizations, a targeted studied is warranted to fully confirm the behavior.

Another limitation of the analysis and annotation scheme presented here is that it only describes the intonational phonology of Middle Variety Urban Shanghainese. As discussed in multiple studies (Tian & Kuang 2019, Gao & Hallé 2017, Takahashi 2013), the language has undergone major sound changes that directly affect many of the generalizations made here, namely the tonal inventory and phrasal accent. In the tonal domain, younger speakers seem to rely more on pitch than phonation cues, which would necessitate more contrastive tones, basically approximating the traditional 5-tone citation tone analysis presented in some studies. Though the extent of difference between younger and older speakers has not been investigated from an intonational view, there are no doubt places where the system as presented here would not be applicable to younger speakers.

6.4. Future Work

Perhaps the clearest needed work is simply the collection and analysis of more data of different kinds to gain a more complete picture of Shanghainese intonation. Specifically, the tone-bearing IP-final boundary tones are in need of exploration. This could be done both with elicitation of short dialogues or with investigation of more naturalistic speech; either could elucidate both the phonological behavior and pragmatic uses of each boundary tone. Post-focal effects are also of interest, as there yet remain questions about the application of de-accenting; again, annotating and analyzing more naturalistic data would be crucial to forming a full understanding of the strategies' use.

Also, as with any full ToBI model, there must be a community of researchers that help codify best practices for labelling. As of yet, the annotation scheme presented here is my work alone, and it would no doubt be helped by researchers from multiple disciplines adding their expertise and knowledge to a future Sh_ToBI system. With such a labelling community formed, there could be testing of inter-labeler reliability of the system, allowing a wide application across multiple fields, including linguistic corpus work, discourse analysis, applied linguistics, as well as theoretical phonology and its interfaces with phonetics and syntax. Another potential route of development would be to add a specifically phonetic tier or representation to Sh_ToBI labelling, making clear the two levels that the tones seem to exist on. Such a change could also allow for more research specifically investigating the usefulness of either characterization of the language.

As mentioned throughout this dissertation, the generational differences among Shanghainese speakers can be stark; this makes the study of younger speakers' intonation all the more interesting. While the differences have been approached from phonetic standpoint (e.g. the loss of breathiness in favor of pitch distinctions and the early inserted low tone targets), it would

be enlightening to see a phonological analysis of their intonational systems done in the same autosegmental-metrical framework, allowing for a comparison of the systems as a whole. Such a study could also begin to shed light on sound change as it applies within the realm of intonation and prosody—a topic that has yet to receive much attention.

Lastly, while many studies have explained how low-level phrasing can be affected by various factors, like footing and morphological structure (Duanmu 1995), syntax (Selkirk & Shen 1990), and frequency and semantic transparency (Yan 2018), no account has tried to provide a predictive account of phrasing that also accounts for the variation seen in production. With the advent of variable approaches to phonological grammar, like the Gradual Learning Algorithm (Boersma 1997) and maximum entropy (Goldwater & Johnson 2003), it has become possible to account for and predict non-categorical processes, whatever they may be. Maximum entropy constraint-based grammars have already been developed to predict variable phrasing patterns, alone as in Yoon (2007) for English, or as part of a larger mathematical structure as in Li et al. (2004) for Mandarin. Due to the relatively large corpus of syntactically controlled data collected in the most recent recordings for this dissertation, such an approach seems natural for Shanghainese as well. While it may be difficult to include all effects on phrasing in one single grammar, having an analysis that is able to correctly balance the syntactic and phonological constraints of Shanghainese phrasing would be a step towards a better understanding of the language.

Appendix A: Speaker Information

This appendix includes the collected demographic information of all speakers recorded in the three datasets introduced in Chapter 3, given in two tables. The first, Table 16 includes: speaker ID, birth year, age at time of recording, age in 2020, gender, and the main district(s) in Shanghai where they grew up and/or live(d). The second table, Table 17, repeats the speaker ID and district information, but additionally shows the results of the dialect survey, which was made from 12 pairs of syllables/characters (taken from Z. Chen 1999) that have or have not undergone mergers in various Shanghainese subdialects. The speakers self-reported their perceived pronunciation of the syllables, with a **Y**es response indicating a merged pair, and a **N**o response indicating a minimal pair.

In both tables, bolded unshaded cells indicate speakers that are included in the analysis above, while unbolded shaded cells show speakers that were not analyzed due to various reasons. These were most often due to difficulty in fluent production on the part of the speaker, or bad audio quality of certain recordings, though a few were excluded due to their low number of mergers (when there were less than seven mergers). Speakers' ID numbers are a number followed by a, b, or c, which corresponds to Datasets I, II, and III respectively. Home districts are listed with the abbreviations given below:

a) Changning CN
b) Chongming CM
c) Hongkou HK
d) Huangpu HP
e) Jing'an JA
f) Luwan LW
g) Pudong PD
h) Yangpu YP
i) Xuhui XH

Speaker ID	Birth Year	Recording Age	Age in 2020	Gender	District
1a	**1964**	**51**	**56**	**F**	**CN, HP**
2a	1955	60	65	F	JA, HP
3a	1966	49	54	F	LW
4a	**1945**	**70**	**75**	**F**	**CN**
5a	**1963**	**52**	**57**	**F**	**CN**
6a	1949	66	71	M	HP
7a	**1961**	**54**	**59**	**F**	**HP**
8a	1958	57	62	F	XH
9a	**1946**	**69**	**74**	**F**	**PD**
10a	**1950**	**65**	**70**	**M**	**HP**
1b	**1966**	**51**	**54**	**F**	**XH**
2b	**1975**	**42**	**45**	**F**	**LW**
3b	1995	22	25	F	XH
4b	**1970**	**47**	**50**	**F**	**XH**
5b	**1957**	**60**	**63**	**F**	**XH, JA**
6b	**1966**	**51**	**54**	**M**	**CM, HP, HK**
1c	**1970**	**49**	**50**	**M**	**HK**
2c	**1963**	**56**	**57**	**M**	**JA**
3c	**1970**	**49**	**50**	**F**	**YP**
4c	1963	56	57	M	HP
5c	1959	60	61	M	YP
6c	**1970**	**49**	**50**	**F**	**HK**
7c	**1974**	**45**	**46**	**F**	**HK**
8c	**1937**	**82**	**83**	**M**	**PD**
9c	**1974**	**45**	**46**	**F**	**HP**
10c	1943	76	77	F	HP
11c	1969	50	51	F	YP
12c	1966	53	54	M	JA
13c	**1968**	**51**	**52**	**M**	**HP**
14c	1962	57	58	F	HP
15c	**1962**	**57**	**58**	**F**	**HP, CN**
16c	1961	58	59	F	HP
17c	1959	60	61	F	LW
Average	**1962.3**	**55.1**	**57.7**		

Table 16. Summary table of all speakers, with speaker ID, birth year, age at recording, age as of 2020, gender, main and main home district(s) in Shanghai. Bolded speakers were included in the analysis for this dissertation. See previous page (p. 150) for district abbreviations.

Speaker ID	District	书 = 诗	来 = 兰	托 = 秃	住 = 树	衣 = 烟	动 = 洞	懂 = 冻	河 = 湖	铜 = 动	夫 = 呼	雷 = 来	冯 = 红	Total # of Mergers
1a	**CN, HP**	**Y**	**Y**	**Y**	**Y**	**Y**	**Y**	**Y**	**Y**	**Y**	**N**	**N**	**N**	**9**
2a	JA, HP	Y	Y	Y	Y	Y	N	Y	N	Y	N	N	N	7
3a	LW	Y	N	Y	Y	Y	Y	N	Y	N	Y	N	N	7
4a	**CN**	**Y**	**Y**	**Y**	**Y**	**Y**	**Y**	**Y**	**Y**	**N**	**N**	**N**	**N**	**8**
5a	**CN**	**Y**	**Y**	**Y**	**Y**	**Y**	**N**	**Y**	**Y**	**Y**	**N**	**Y**	**N**	**9**
6a	HP	Y	Y	Y	Y	Y	Y	N	N	N	N	N	N	6
7a	**HP**	**N**	**Y**	**Y**	**Y**	**N**	**Y**	**Y**	**Y**	**Y**	**N**	**N**	**N**	**7**
8a	XH	Y	Y	N	N	N	N	Y	Y	N	N	N	N	4
9a	**PD**	**Y**	**Y**	**Y**	**Y**	**Y**	**Y**	**Y**	**Y**	**Y**	**Y**	**Y**	**N**	**11**
10a	**HP**	**Y**	**Y**	**Y**	**Y**	**N**	**Y**	**Y**	**N**	**Y**	**Y**	**N**	**N**	**8**
1b	**XH**	**Y**	**Y**	**Y**	**Y**	**Y**	**N**	**N**	**Y**	**N**	**N**	**N**	**N**	**6**
2b	**LW**	**Y**	**Y**	**Y**	**Y**	**Y**	**Y**	**N**	**Y**	**N**	**N**	**Y**	**N**	**8**
3b	XH	Y	Y	Y	Y	Y	Y	Y	N	N	N	N	N	7
4b	**XH**	**Y**	**Y**	**Y**	**Y**	**Y**	**Y**	**Y**	**Y**	**Y**	**Y**	**N**	**N**	**10**
5b	**XH, JA**	**Y**	**Y**	**Y**	**Y**	**Y**	**Y**	**N**	**Y**	**Y**	**N**	**N**	**N**	**8**
6b	**CM, HP, HK**	**Y**	**Y**	**Y**	**N**	**Y**	**Y**	**Y**	**Y**	**Y**	**N**	**Y**	**N**	**9**
1c	**HK**	**Y**	**Y**	**N**	**Y**	**Y**	**Y**	**Y**	**Y**	**N**	**N**	**Y**	**N**	**8**
2c	**JA**	**Y**	**Y**	**Y**	**Y**	**Y**	**Y**	**Y**	**Y**	**Y**	**Y**	**Y**	**N**	**11**
3c	**YP**	**Y**	**Y**	**Y**	**Y**	**Y**	**Y**	**Y**	**Y**	**Y**	**Y**	**Y**	**N**	**11**
4c	HP	Y	Y	Y	Y	Y	Y	Y	N	Y	N	N	N	8
5c	YP	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	N	11
6c	**HK**	**Y**	**Y**	**Y**	**Y**	**Y**	**Y**	**Y**	**Y**	**Y**	**Y**	**Y**	**N**	**11**
7c	**HK**	**Y**	**Y**	**Y**	**Y**	**Y**	**Y**	**Y**	**Y**	**Y**	**N**	**N**	**N**	**9**
8c	**PD**	**Y**	**Y**	**Y**	**Y**	**Y**	**Y**	**Y**	**Y**	**Y**	**Y**	**Y**	**N**	**11**
9c	**HP**	**Y**	**Y**	**Y**	**Y**	**Y**	**Y**	**Y**	**Y**	**Y**	**Y**	**Y**	**N**	**11**
10c	HP	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	N	11
11c	YP	Y	Y	Y	N	Y	Y	Y	Y	Y	Y	N	N	9
12c	JA	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	N	11
13c	**HP**	**Y**	**Y**	**Y**	**Y**	**Y**	**Y**	**Y**	**Y**	**Y**	**N**	**Y**	**N**	**10**
14c	HP	Y	Y	Y	Y	Y	Y	N	Y	Y	Y	N	N	9
15c	**HP, CN**	**Y**	**Y**	**Y**	**Y**	**Y**	**Y**	**Y**	**N**	**Y**	**N**	**N**	**N**	**8**
16c	HP	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	N	11
17c	LW	N	N	Y	N	N	N	Y	N	N	N	N	N	2
# of spkrs w/ merger:		**31**	**31**	**31**	**29**	**29**	**28**	**27**	**26**	**24**	**15**	**15**	**0**	**Average # of Mergers: 8.67**

Table 17. Summary of dialect survey with 12 pairs of syllables/characters, along with Speaker ID and home district(s). See p. 150 above for district abbreviations. 'Y' indicates a self-response of merger, while 'N' indicates a lack of merger for the indicated pair. Bolded speakers were included in the analysis for this dissertation.

Appendix B: Stimuli

This appendix includes all stimuli presented to speakers, broken up into the three datasets that were outlined in Chapter 3. Full descriptions of the factors influencing stimuli creation are given there. The lists below specifically include the following information: stimuli as presented to the speakers, in Simplified Chinese characters; a broad surface transcription of the stimuli in IPA, with every syllable followed by its traditional citation tone number (T1-5); and the idiomatic English translation of the stimuli.

Dataset I Stimuli

	Simplified Characters	**IPA**	**English Translation**
1	编排	[pi1 ba3]	to arrange
2	报到	[pɔ2 dɔ3]	to report
3	另外	[liɲ3 ŋɑ3]	another
4	一样	[iəʔ4 ɦiã]	same
5	鼻头	[biəʔ5 dɤ3]	nose
6	香皮皂	[ɕiã1 bi3 zɔ3]	scented soap
7	草棚棚	[tsʰɔ2 bã3 bã3]	thatched hut
8	被横头	[bi3 ɦuã3 dɤ3]	comforter-protecting fabric; translated as "comforter*" below
9	雪里蕻	[ɕyøʔ4 li3 ɦoŋ3]	potherb mustard
10	白醪醪	[baʔ5 liɔ3 liɔ3]	sickly
11	三日两头	[sɛ1 ɲiəʔ5 liã3 dɤ3]	always
12	要样无样	[iɔ2 ɦiã3 m3 ɦiã3]	can't tell
13	无头无脑	[m3 dɤ3 m3 nɔ3]	idiotic
14	一时上头	[iəʔ4 zz3 zã3 dɤ3]	in a short time
15	热门热路	[ɲiəʔ5 məŋ ɲiəʔ5 lu3]	hot; popular
16	花无百日红	[ho1 m3 paʔ4 ɲiəʔ5 ɦoŋ3]	"looks fade"
17	开空头支票	[kʰe2 koŋ2 dɤ3 tsz2 pʰiɔ2]	"make a false promise"
18	远开八只脚	[y3 kʰe2 paʔ4 tsəʔ4 tɕiaʔ4]	"very far away"
19	老实勿客气	[lɔ3 zz3 vəʔ5 kʰaʔ4 tɕʰi2]	"to be honest with you"
20	老是勿客气	[lɔ3 zz3 vəʔ5 kʰaʔ4 tɕʰi2]	"always impolite"
21	一场无结果	[iəʔ tsã1 m3 tɕiəʔ4 ku2]	"do something in vain"
22	日光搭火光	[ɲiəʔ5 kuã1 tʰaʔ4 hu2 kuã1]	"use electricity while it's still light outside"

	Simplified Characters	IPA	English Translation
23	___是伊讲的	…[zz3 ɦi3 kã ɣəʔ5]	___ is what she said
24	伊讲___给侬听	[ɦi3 kã]…[pəʔ4 noŋ3 tʰiɲ1]	He said ___ for you to hear
25	伊讲的是___	[ɦi3 kã ɣəʔ5 zz3]…	What she said was ___
26	伊讲的是___？！	[ɦi3 kã ɣəʔ5 zz3]…?!	What he said was ___?! (echo question / disbelief)
27	这只节目是伊编排的。	[kəʔ5 tsəʔ4 tɕiəʔ4 moʔ5 zz3 ɦi3 pi1 ba3 ɣəʔ]	This program was arranged by her
28	伊对草莓过敏	[ɦi3 tɛ2 tsʰɔ2 me3 ku2 miɲ3]	He's allergic to strawberries
29	伊拉屋里前头种了较关草莓	[ɦi3 la3 oʔ4 li3 tɕʰi2 dɤ3 tsoŋ2 ləʔ5 tɕiɔ1 kue1 tsʰɔ2 me3]	There are many strawberries growing in front of their house
30	草莓落了这个地方长得勿好	[tsʰɔ2 me3 laʔ5 laʔ5 gəʔ5 ɣəʔ5 di3 fã1 tsã2 təʔ4 vəʔ5 hɔ2]	Strawberries don't grow well in this place
31	爸爸昨日找到一只大的青蛙	[pa1 pa1 soʔ4 ɲiəʔ5 tsɔ2 tɔ2 iəʔ4 tsəʔ4 tu1 ɣəʔ5 tɕʰiɲ1 o2]	Dad found a big frog yesterday
32	侬是礼拜三来报到的伐？	[noŋ3 zz3 li3 ba3 sɛ1 le3 pɔ2 dɔ3 ɣəʔ5 va3]	Did you report on Wednesday?
33	另外一个原因就是我根本没有辰光	[liɲ3 ŋɑ3 iəʔ4 ɣəʔ5 ɲy3 iɲ1 dʑiɤ3 zz3 ŋu3 kəŋ1 pəŋ2 m3 məʔ5 zəŋ3 kuã1]	Another reason is I simply have no time
34	侬想另外来上课伐？	[noŋ3 ɕiã liɲ3 ŋɑ3 le3 zã kʰu2 va3]	Do you want to attend another class?
35	我买的洋伞跟侬的一样的	[ŋu3 ma3 ɣəʔ5 ɦiã3 sã2 kəŋ1 noŋ3 ɣəʔ5 iəʔ4 ɦiã3 ɣəʔ5]	The umbrella I bought is the same as yours
36	伊拉两个人上一样的课	[ɦi3 la3 liã3 ɣəʔ5 ɲiɲ3 zã iəʔ4 ɦiã3 ɣəʔ5 kʰu2]	Those two people take the same class
37	达个两根针一样粗细	[taʔ4 ɣəʔ5 liã3 kəŋ1 tsəŋ1 iəʔ4 ɦiã3 tsʰu1 ɕi2]	Those two needles are the same thickness
38	鼻头疼死啦!	[biəʔ3 dɤ3 təŋ2 sz2 laʔ5]	My nose hurts!
39	伊突然打坏忒了我的鼻头	[ɦi3 təʔ4 zø3 dã3 ɦua3 tʰəʔ4 ləʔ5 ŋu3 ɣəʔ5 biəʔ3 dɤ3]	She suddenly broke my nose.
40	伊的鼻头老高的	[ɦi3 ɣəʔ5 biəʔ3 dɤ3 lɔ3 kɔ1 ɣəʔ5]	His nose is long
41	我让伊去买两块香皮皂	[ŋu3 ɲiã ɦi3 tɕʰi2 ma3 liã3 kʰue2 ɕiã1 bi3 zɔ3]	I sent him to go buy two pieces of scented soap.
42	这块香皮皂味道老好闻咯	[gəʔ5 kʰue2 ɕiã1 bi3 zɔ3 mi3 dɔ3 lɔ3 hɔ2 ɣəʔ5]	This scented soap smells great
43	香皮皂对侬皮肤好	[ɕiã1 bi3 zɔ3 te2 noŋ3 bi3 fu2 hɔ2]	Scented soap does your skin good
44	这只房子就是草棚棚	[gəʔ5 kʰue2 vã3 tsz2 dʑiɤ3 zz3 tsʰɔ2 bã3 bã3]	This house is just a hut

	Simplified Characters	IPA	English Translation
45	城市里厢的草棚棚老可惜咯	[tsəŋ2 sz2 li3 ɕiã ɣəʔ5 tsʰɔ2 bã3 bã3 lɔ3 kʰo2 ɕi2 ɣəʔ5]	The huts in the city are pitiful
46	草棚棚着火了	[tsʰɔ2 bã3 bã3 zaʔ5 ho2 ləʔ5]	The hut's on fire
47	我现在无没辰光汰我的被横头	[ŋu3 ɦi3 ze3 m3 məʔ55 zəŋ3 kuã1 da3 ŋu3 ɣəʔ5 bi3 ɦuã3 dɤ3]	I don't have time right now to wash my comforter*
48	侬的被横头老喔龌咯	[noŋ3 ɣəʔ5 bi3 ɦuã3 dɤ3 lɔ3 oʔ4 tsʰoʔ4 ɣəʔ5]	Your comforter* is dirty
49	被横头应该汰一汰	[bi3 ɦuã3 dɤ3 iɲ1 ke1 da3 iəʔ4 da3]	The comforter* should be washed
50	这个菜是雪里蕻炒豆瓣	[gəʔ5 ɣəʔ5 tsʰe2 zz3 ɕyøʔ4 li3 ɦoŋ3 tsʰɔ3 dɤ3 be3]	This dish is potherb mustard fried with beans
51	雪里蕻是我顶欢喜的菜之一	[ɕyøʔ4 li3 ɦoŋ3 zz3 ŋu3 tiɲ1 huø1 ɕi2 ɣəʔ5 tsʰe2 tsəʔ4 iəʔ4]	Potherb mustard is one of my favorite vegetables
52	阿拉应该等油热了才放雪里蕻	[aʔ4 la3 iɲ1 ke1 təŋ2 ɦiɤ3 ɲiəʔ5 ləʔ5 tsʰe2 fã2 ɕyøʔ4 li3 ɦoŋ3]	We should wait until the oil is hot to put in the potherb mustard
53	侬的面色白醪醪了	[noŋ3 ɣəʔ5 mi3 səʔ4 baʔ5 liɔ3 liɔ3 ɣəʔ5]	Your face looks sickly
54	白醪醪的小人勿欢喜上体育课	[baʔ5 liɔ3 liɔ3 ɣəʔ5 ɕiɔ2 ɲiɲ3 vəʔ5 huø1 ɕi2 zã tʰi2 ioʔ kʰu2]	Sickly children don't like to go to PE class
55	医生到伊屋里厢帮白醪醪的小人看毛病	[i1 sã1 tɔ2 ɦi3 oʔ4 li3 ɕiã2 pã baʔ5 liɔ3 liɔ3 ɣəʔ5 ɕiɔ2 ɲiɲ3 kʰø2 mɔ3 biɲ3]	The doctor came to his house to help the sickly child with his sickness.
56	椅子高头有三只橘子。	[ɦi3 tsz2 kɔ1 dɤ3 ɦiɤ3 sɛ1 tsəʔ4 tɕyøʔ4 tsz2]	There are three oranges on the chair
57	箱子上头有一杯绿茶	[ɕiã1 tsz2 zã3 dɤ3 ɦiɤ3 iəʔ4 pe1 loʔ5 zo3]	There is a cup of green tea on the box
58	余太太明朝去看三间老房子	[ɦy3 tʰa2 tʰa2 miɲ3 tsɔ2 tɕʰi2 kʰø2 sɛ1 ge3 lɔ3 vã3 tsz2]	Mrs. Yu is going to see three old houses tomorrow
59	落了上海的地皮越来越少，所以地皮老贵咯	[laʔ4 laʔ4 zã3 he2 ɣəʔ5 di3 bi3 yøʔ4 le3 yøʔ4 sɔ2 | so2 i2 di3 bi3 lɔ3 dʑy3 ɣəʔ5]	Real estate in Shanghai is rarer and rarer, so real estate is expensive
60	因为周围山多，所以没有可以用的地皮	[iɲ1 ue2 tsɤ1 ue2 sɛ1 tu1 | so2 i2 m3 məʔ4 kʰo2 i2 ɦyoŋ3 ɣəʔ5 di3 bi3]	Since the surroundings are mountainous, there's not any land to be used.
61	我现落在住了加里佛尼亚	[ŋu3 ɦi3 ze3 laʔ5 laʔ5 ka1 li3 foʔ4 ɲi3 ia2]	I live in California now.
62	加里佛尼亚是美国的州之一	[ka1 li3 foʔ4 ɲi3 ia2 zz3 me3 koʔ4 ɣəʔ5 tsɤ1 tsz2 iəʔ4]	California is one of the US's states

	Simplified Characters	IPA	English Translation
64	阿拉明朝去上海博物馆好伐？	[aʔ4 la3 miɲ3 tsɔ2 tɕʰi2 zã3 he2 bəʔ vəʔ5 ku2 hɔ2 va3]	Let's go to the Shanghai Museum tomorrow, all right?
65	余先生关照爷爷今年春天应该老热的	[ɦy3 ɕi1 sã1 kue1 tsɔ2 ia2 ia2 tɕiɲ1 ɲi3 tsəŋ1 tʰi1 iɲ1 ke1 lɔ3 ɲiəʔ5 lə5]	Mr. Yu told grandpa that spring should be hot this year
66	伊关照爸爸椅子坏忒了	[ɦi3 kue1 tsɔ2 pa1 pa1 ɦi3 tsz2 ɦua3 tʰəʔ4 ləʔ4]	She told dad that the chair broke
67	爸爸勿要姆妈的礼物	[pa1 pa1 vəʔ5 iɔ2 m1 ma2 ɣəʔ5 li3 vəʔ5]	Dad doesn't want mom's gift
68	阿拉姆妈炒的虾仁比啥人炒的还要好吃	[aʔ4 la3 m1 ma3 tsʰɔ2 ɣəʔ5 hø1 ɲiɲ3 pi2 sa2 ɲiɲ3 ɣəʔ5 e2 iɔ2 hɔ2 tɕʰiəʔ4]	Our mom cooks shrimp better than anyone
69	伊落在饭店里点了虾仁炒蛋	[ɦi3 laʔ5 laʔ5 ve3 ti2 li3 ti2 ləʔ5 hø1 ɲiɲ3 tsʰɔ2 dɛ3]	He ordered shrimp fried egg at the restaurant
70	虾仁炒蛋是这个城市的特色菜	[hø1 ɲiɲ3 tsʰɔ2 dɛ3 zz3 gəʔ5 ɣəʔ5 tsəŋ2 sz2 ɣəʔ5 dəʔ5 səʔ4 tsʰe2]	Shrimp fried egg is this city's specialty
71	老师听说侬还无没做功课	[lɔ3 sz2 tʰiɲ1 səʔ4 noŋ3 e2 m3 məʔ5 tsu2 koŋ1 kʰu2]	The teacher heard you still haven't done your homework
72	我今朝到苏州去	[ŋu3 tɕiɲ1 tsɔ2 tɔ2 su1 tsɤ1 tɕʰi2]	I'm going to Suzhou today
73	伊有一只高的椅子	[ɦi3 ɦiɤ3 iəʔ4 tsəʔ4 kɔ1 ɣəʔ5 ɦi3 tsz2]	He has a big chair
74	我今朝到超市去买西瓜，萝卜，茶叶，嫩豆腐，香瓜子，水果糖，辣油，交关东西。	[ŋu3 tɕiɲ1 tsɔ2 tɔ2 tsʰɔ2 sz2 tɕʰi2 ma3 ɕi1 ku1 | lu3 boʔ5 | zo3 iəʔ4 | nəŋ3 dɤ3 vu3 | ɕiã1 ku1 tsz2 | sz2 ku2 dã3 | laʔ5 ɦiɤ3 | tɕiɔ1 kue1 məʔ5 zz3]	Today, I'm going to the supermarket to buy watermelon, carrots, tea, soft tofu, fragrant melon seeds, fruit sugar, spicy oil, and some other stuff
75	椅子上头有张报纸	[ɦi3 tsz2 zã dɤ3 ɦiɤ3 tsã2 pɔ2 tsz2]	There's a newspaper on the chair
76	红的椅子上头有张报纸	[ɦoŋ3 ɣəʔ5 ɦi3 tsz2 zã dɤ3 ɦiɤ3 tsã2 pɔ2 tsz2]	There's a newspaper on the red chair
77	高的红的椅子上头有张报纸	[kɔ1 ɣəʔ5 ɦoŋ3 ɣəʔ5 ɦi3 tsz2 zã dɤ3 ɦiɤ3 tsã2 pɔ2 tsz2]	There's a newspaper on the big red chair
78	贵的高的红的椅子上头有张报纸	[dʑy3 ɣəʔ5 kɔ1 ɣəʔ5 ɦoŋ3 ɣəʔ5 ɦi3 tsz2 zã dɤ3 ɦiɤ3 tsã2 pɔ2 tsz2]	There's a newspaper on the expensive big red chair
79	余先生的贵的高的红的椅子上头有张报纸	[ɦy3 ɕi1 sã1 ɣəʔ5 dʑy3 ɣəʔ5 kɔ1 ɣəʔ5 ɦoŋ3 ɣəʔ5 ɦi3 tsz2 zã dɤ3 ɦiɤ3 tsã2 pɔ2 tsz2]	There's a newspaper on Mr. Yu's expensive big red chair
80	A: 余先生关照奶奶伐？	[ɦy3 ɕi1 sã1 kue1 tsɔ2 ne3 ne3 va3]	A: Did Mr. Yu tell grandma?

	Simplified Characters	IPA	English Translation
	B: 勿是咯，余先生是关照**爷爷**咯。	[vəʔ5 zz3 ɣəʔ5 | ɦy3 ɕi1 sã1 zz3 kue1 tsɔ2 ia2 ia2 ɣəʔ5]	B: No, Mr. Yu told **grandpa**
81	A: 侬晓得余太太今朝做啥？	[noŋ3 ɕiɔ2 təʔ4 ɦy3 tʰa2 tʰa2 tɕiɲ1 tsɔ2 tsu2 sa2]	A: Do you know what Mrs. Yu is doing today?
	B: 勿晓得，但是余太太**明朝**要去看点老房子。	[vəʔ5 ɕiɔ2 təʔ4 | de3 zz3 ŋu3 ɕiɔ2 təʔ4 ɦy3 tʰa2 tʰa2 miɲ3 tsɔ2 iɔ2 tɕʰi2 ti2 lɔ3 vã3 tsz2]	B: No, but **tomorrow** she's going to see a few old houses.
82	A: 绿茶是伊顶欢喜的一种茶。我一点都勿欢喜。	[loʔ5 zo3 zz3 ɦi3 tiɲ1 huø1 ɕi2 ɣəʔ5 iəʔ4 tsoŋ2 zo3 || ŋu3 iəʔ4 ti2 tu1 vəʔ5 huø1 ɕi2]	A: Green tea is his favorite kind of tea. I don't like it one bit.
	B: 侬勿欢喜绿茶？！侬为啥勿欢喜绿茶？	[noŋ3 vəʔ5 huø1 ɕi2 loʔ5 zo3 || noŋ3 ue2 sa2 vəʔ5 huø1 ɕi2 loʔ5 zo3]	B: You don't like green tea?! Why don't you like green tea?
83	A: 我老欢喜小动物咯。昨日看到了一只特别可爱的小狗。	[ŋu3 lɔ3 huø1 ɕi2 ɕiɔ2 doŋ3 vəʔ5 ɣəʔ5 || soʔ4 ɲiəʔ5 kʰø2 tɔ2 ləʔ5 iəʔ4 tsəʔ4 dəʔ5 biəʔ5 kʰo2 e2 ɣəʔ5 ɕiɔ2 kɤ2]	A: I really like little animals. Yesterday I saw an especially cute little dog
	B: 听勿清爽，是小猫伐？	[tʰiɲ1 vəʔ5 tɕʰiɲ1 sã1 | zz3 ɕiɔ2 mɔ1 va3]	B: I didn't catch that; was it a kitten?
	A: 不是，是小狗。	[vəʔ5 zz3 | zz3 ɕiɔ2 kɤ2]	A: No, it was a little dog.
84	A: 爸爸欢喜姆妈的礼物伐？	[pa1 pa1 huø1 ɕi2 m1 ma3 ɣəʔ5 li3 vəʔ5 va3]	A: Did dad like mom's gift?
	B: 勿是咯，爸爸**勿欢喜**姆妈的礼物。	[vəʔ zz3 ɣəʔ5 | pa1 pa1 vəʔ5 huø1 ɕi2 m1 ma3 ɣəʔ5 li3 vəʔ5]	B: No, dad **refused** mom's gift.
85	A: 箱子里头有勿有茶叶？	[ɕiã1 tsz2 li3 dɤ3 ɦiɤ3 vəʔ5 ɦiɤ3 zo3 iəʔ4]	A: Are there tea leaves in the box?
	B: 箱子**高头**有一罐茶叶。	[ɕiã1 tsz2 kɔ1 dɤ3 ɦiɤ iəʔ kuø2 zo3 iəʔ4	B: There's a tin of tea leaves **on** the box.

Dataset II Stimuli

	Simplified Characters	IPA	English Translation
86	妹妹上年买了内**衣**	[me3 me3 zã3 ɲi3 ma3 ləʔ5 ne3 i1]	Little sister bought underwear last year.
87	妹妹认为拉面顶好	[me3 me3 ɲiɲ3 ue2 la1 mi3 tiɲ1 hɔ2]	Little sister thinks ramen is the best.
88	阿姨认为澳门老好	[a2 ɦi3 ɲiɲ2 ɦue3 ɔ2 məŋ3 lɔ3 hɔ2]	Auntie thinks Macau is great.

	Simplified Characters	IPA	English Translation
89	妹妹认为银行太远了	[me3 me3 ɲiɲ ɦue3 ɦiɲ3 ɦã3 tʰa2 ɦy3 ləʔ5]	Little sister thinks the bank is too far.
90	姆妈今年安排婚礼	[m1 ma2 tɕiɲ1 ɲi3 ø1 ba3 hun1 li3]	Mom is planning a wedding this year.
91	姆妈刚刚买了饮料	[m1 ma3 kã1 kã1 ma3 ləʔ5 iɲ2 liɔ3]	Mom just bought some beverages.
92	姆妈要求增加工资	[m1 ma3 iɔ1 dʑiɤ3 tsəŋ1 ka1 koŋ1 tsz2]	Mom requested a pay raise.
93	小狗好像欢喜虾仁	[ɕiɔ2 kɤ2 hɔ2 ɕiã2 huø1 ɕi2 hø1 ɲiɲ3]	The puppy seems to like shrimp.
94	阿姨要求妹妹运动	[a2 ɦi3 iɔ1 dʑiɤ3 me3 me3 ɦyɲ3 doŋ3]	Auntie requests that little sister exercise.
95	姆妈欢喜东京的拉面	[m1 ma3 huø1 ɕi2 toŋ1 tɕiɲ1 ɣəʔ5 la1 mi3]	Mom likes Tokyo ramen.
96	妹妹已经关照姆妈了	[me3 me3 i2 tɕiɲ1 kue1 tsɔ2 m1 ma3 ləʔ5]	Little sister already told mom.
98	妹妹马路上买了油条	[me3 me3 mo3 lu3 lã3 ma3 ləʔ5 ɦiɤ3 diɔ3]	Little sister bought a donut on the street.
99	姆妈安排明朝的晚会	[m1 ma3 ø1 ba3 miɲ3 tsɔ2 ɣəʔ5 ue2 ue2]	Mom is planning tomorrows party.
100	伊的美容院门面老好咯	[ɦi3 ɣəʔ5 me3 ɦoŋ3 ɦy3 məŋ3 mi3 lɔ3 hɔ2 ɣəʔ5]	Her beauty parlor’s façade is really nice.
101	这家便利店饮料老多咯	[gəʔ5 ka2 bi3 li3 ti2 iɲ2 liɔ3 lɔ3 tu1 ɣəʔ5]	This convenience store has a lot of beverages.
102	今朝姆妈的围裙老龌龊咯	[tɕiɲ1 tsɔ2 m1 ma3 ɣəʔ5 ue2 dʑyɲ3 lɔ3 oʔ4 tsʰoʔ ɣəʔ5]	Today mom’s apron is dirty.
103	姆妈的香肥皂味道老好咯	[m1 ma3 ɣəʔ5 ɕiã1 bi3 zɔ3 mi3 dɔ3 lɔ3 hɔ2 ɣəʔ5]	Mom’s scented soap—the smell is so good.

Dataset III Stimuli

	Simplified Characters	IPA	English Translation
104	伊骂我	[ɦi3 mo3 ŋu3]	He scolds me.
105	小李骂我	[ɕiɔ2 li3 mo3 ŋu3]	Little Lie scolds me.
106	董事长骂我	[tɔŋ2 zz3 tsã2 mo3 ŋu3]	The trustee scolds me.
107	女朋友骂我	[ɲy3 bəŋ3 ɦiɤ3 mo3 ŋu3]	The girlfriend scolds me.
108	苏州校长骂我	[su1 tsɤ2 ɕiɔ2 tsã2 mo3 ŋu3]	The Suzhou principal scolds me.
109	伊骂小吴	[ɦi3 mo3 ɕiɔ2 u1]	He scolds Little Wu.
110	小李骂小吴	[ɕiɔ2 li3 mo3 ɕiɔ2 u1]	Little Li scolds Little Wu.
111	董事长骂小吴	[tɔŋ2 zz3 tsã2 mo3 ɕiɔ2 u1]	The trustee scolds Little Wu.

	Simplified Characters	IPA	English Translation
112	女朋友骂小吴	[ɲy3 bəŋ3 ɦiɤ3 mo3 ɕiɔ2 u1]	The girlfriend scolds Little Wu.
113	苏州校长骂小吴	[su1 tsɤ2 ɕiɔ2 tsɑ̃2 mo3 ɕiɔ2 u1]	The Suzhou principal scolds Little Wu.
114	搿眼人买油条	[keʔ4 ɲe3 ɲiɲ3 ma3 ɦiɤ3 diɔ3]	These few people buy donuts.
115	五个人买油条	[ŋ3 ɣəʔ5 ɲiɲ3 ma3 ɦiɤ3 diɔ3]	Five people buy donuts.
116	搿五个人买油条	[kəʔ4 ŋ3 ɣəʔ5 ɲiɲ3 ma3 ɦiɤ3 diɔ3]	These five people buy donuts.
117	十五个人买油条	[zəʔ5 ŋ3 ɣəʔ5 ɲiɲ3 ma3 ɦiɤ3 diɔ3]	Fifteen people buy donuts.
118	搿十五个人买油条	[taʔ4 zəʔ5 ŋ3 ɣəʔ5 ɲiɲ3 ma3 ɦiɤ3 diɔ3]	These fifteen people buy donuts.
119	搿眼小人买油条	[kəʔ4 ɲe3 ɕiɔ2 ɲiɲ3 ma3 ɦiɤ3 diɔ3]	These few kids buy donuts.
120	五个小人买油条	[ŋ3 ɣəʔ5 ɕiɔ2 ɲiɲ3 ma3 ɦiɤ3 diɔ3]	Five kids buy donuts.
121	搿五个小人买油条	[kəʔ4 ŋ3 ɣəʔ5 ɕiɔ2 ɲiɲ3 ma3 ɦiɤ3 diɔ3]	These five kids buy donuts.
122	十五个小人买油条	[zəʔ5 ŋ3 ɣəʔ5 ɕiɔ2 ɲiɲ3 ma3 ɦiɤ3 diɔ3]	Fifteen kids buy donuts.
123	搿十五个小人买油条	[kəʔ4 zəʔ5 ŋ3 ɣəʔ5 ɕiɔ2 ɲiɲ3 ma3 ɦiɤ3 diɔ3]	These fifteen kids buy donuts.
124	侬爷买油条	[na3 ɦia3 ma3 ɦiɤ3 diɔ3]	Y’all’s dad buys donuts.
125	阿拉爷买油条	[a2 la3 ɦia3 ma3 ɦiɤ3 diɔ3]	Our dad buys donuts.
126	我的猫咬小余	[ŋu3 ɣəʔ5 mɔ1 ŋɔ3 ɕiɔ2 ɦy3]	My cat bites Little Yu.
127	妹妹的猫咬小余	[me3 me3 ɣəʔ5 mɔ1 ŋɔ3 ɕiɔ2 ɦy3]	Little sister’s cat bites Little Yu.
128	女朋友的猫咬小余	[ɲy3 bəŋ3 ɦiɤ3 ɣəʔ5 mɔ1 ŋɔ3 ɕiɔ2 ɦy3]	The girlfriend’s cat bites Little Yu.
129	董事长的猫咬小余	[toŋ2 zz3 tsɑ̃2 ɣəʔ5 mɔ1 ŋɔ3 ɕiɔ2 ɦy3]	The trustee’s cat bites Little Yu.
130	侬姆妈买油条	[na3 m1 ma1 ma3 ɦiɤ3 diɔ3]	Y’all’s mom buys donuts.
131	阿拉姆妈买油条	[a2 la3 m1 ma1 ma3 ɦiɤ3 diɔ3]	Our mom buys donuts.
132	我的爱人买油条	[ŋu3 ɣəʔ5 e2 ɲiɲ3 ma3 ɦiɤ3 diɔ3]	My spouse buys donuts.
133	妹妹的爱人买油条	[me3 me3 ɣəʔ5 e2 ɲiɲ3 ma3 ɦiɤ3 diɔ3]	Little sister’s spouse buys donuts.
134	女朋友的朋友买油条	[ɲy3 bəŋ3 ɦiɤ3 ɣəʔ5 bəŋ3 ɦiɤ3 ma3 ɦiɤ3 diɔ3]	The girlfriend’s friend buys donuts.
135	董事长的爱人买油条	[toŋ2 zz3 tsɑ̃2 ɣəʔ5 e2 ɲiɲ3 ma3 ɦiɤ3 diɔ3]]	The trustee’s spouse buys donuts.

	Simplified Characters	IPA	English Translation
136	侬爷的猫咬小吴	[na3 ɦia3 ɣəʔ5 mɔ1 ŋɔ3 ɕiɔ2 u1]	Y’all’s dad’s cat bites Little Wu.
137	我的猫的脚抓小吴	[ŋu3 ɣəʔ5 mɔ1 ɣəʔ5 tsiaʔ4 tsa1 ɕiɔ2 u1]	My cat’s food scratches Little Wu.
138	伊妹妹的猫咬小吴	[ɦi3 me3 me3 ɣəʔ5 mɔ1 ŋɔ3 ɕiɔ2 u1]	His little sister’s cat bites Little Wu.
139	伊的爱人的猫咬小吴	[ɦi3 ɣəʔ e2 ɲiŋ3 ɣəʔ5 mɔ1 ŋɔ3 ɕiɔ2 u1]	His spouse’s cat bites Little Wu.
140	伊的女朋友的猫咬小吴	[ɦi3 ɣəʔ5 ɲy3 bəŋ3 ɦiɤ3 ɣəʔ5 mɔ1 ŋɔ3 ɕiɔ2 u1]	His girlfriend’s cat bites Little Wu.
141	伊的董事长的猫咬小吴	[ɦi3 ɣəʔ5 toŋ2 zz3 tsã2 ɣəʔ5 mɔ1 ŋɔ3 ɕiɔ2 u1]	His trustee’s cat bites Little Wu.
142	侬爷的朋友买油条	[na3 ɦia3 ɣəʔ5 bəŋ3 ɦiɤ3 ma3 ɦiɤ3 diɔ3]	Y’all’s dad’s friend buys donuts.
143	我的猫的脚爪抓小李	[ŋu3 ɣəʔ5 mɔ1 ɣəʔ5 tsiaʔ4 tsɔ2 tsa1 ɕiɔ2 li3]	My cat’s claw scratches Little Li.
144	伊妹妹的朋友买油条	[ɦi3 me3 me3 ɣəʔ5 bəŋ3 ɦiɤ3 ma3 ɦiɤ3 diɔ3]	His little sister’s friend buys donuts.
145	伊的爱人的朋友买油条	[ɦi3 ɣəʔ5 e2 ɲiŋ3 ɣəʔ5 bəŋ3 ɦiɤ3 ma3 ɦiɤ3 diɔ3]	His spouse’s friend buys donuts.
146	伊的女朋友的朋友买油条	[ɦi3 ɣəʔ5 ɲy3 bəŋ3 ɦiɤ3 ɣəʔ5 bəŋ3 ɦiɤ3 ma3 ɦiɤ3 diɔ3]	His girlfriend’s friend buys donuts.
147	伊的董事长的朋友买油条	[ɦi3 ɣəʔ5 toŋ2 zz3 tsã3 ɣəʔ5 bəŋ3 ɦiɤ3 ma3 ɦiɤ3 diɔ3]	His trustee’s friend buys donuts.
148	淡的菜对侬好	[de3 ɣəʔ5 tsʰe2 te2 noŋ3 hɔ2]	Light food does you good.
149	健康的人吃蔬菜	[dʑi3 kʰã1 ɣəʔ5 ɲiŋ3 tɕiɪʔ4 su2 tsʰe2]	Healthy people each veggies.
150	毛茸茸的猫咬小李	[mɔ3 ɲioŋ3 ɲioŋ3 ɣəʔ5 mɔ1 ŋɔ3 ɕiɔ2 li3]	The fluffy cat bites Little Li.
151	老淡的菜对侬好	[lɔ3 de3 ɣəʔ5 tsʰe2 te2 noŋ3 hɔ2]	Very light food does you good.
152	老健康的人吃蔬菜	[lɔ3 dʑi3 kʰã1 ɣəʔ5 ɲiŋ3 tɕiɪʔ4 su2 tsʰe2]	Very healthy people eat veggies.
153	淡的小菜对侬好	[de3 ɕiɔ2 ɣəʔ5 tsʰe2 te2 noŋ3 hɔ2]	Light side dishes do you good.
154	健康的小人吃蔬菜	[dʑi3 kʰã1 ɣəʔ5 ɕiɔ2 ɲiŋ3 tɕiɪʔ4 su2 tsʰe2]	Healthy kids each veggies.
155	毛茸茸的小猫咬小李	[mɔ3 ɲioŋ3 ɲioŋ3 ɣəʔ5 ɕiɔ2 mɔ1 ŋɔ3 ɕiɔ2 li3]	The fluffy kitten bites Little Li.
156	老淡的小菜对侬好	[lɔ3 de3 ɕiɔ2 ɣəʔ5 tsʰe2 te2 noŋ3 hɔ2]	Very light side dishes do you good.
157	老健康的小人吃蔬菜	[lɔ3 dʑi3 kʰã1 ɣəʔ5 ɕiɔ2 ɲiŋ3 tɕiɪʔ4 su2 tsʰe2]	Very healthy kids each veggies.
158	我想的人来上海	[ŋu3 ɕiã2 ɣəʔ5 ɲiŋ3 le3 zã3 he2]	The person I miss is coming to Shanghai.

	Simplified Characters	IPA	English Translation
159	我认得的人来上海	[ŋu3 ɕiã2 ɣəʔ5 ɲiɲ3 le3 zã3 he2]	The person I know is coming to Shanghai.
160	我老欢喜的人来上海	[ŋu3 lɔ3 huø1 ɕi2 ɣəʔ5 ɲiɲ3 le3 zã3 he2]	The person I really like is coming to Shanghai.
161	我昨日碰到的人来上海	[ŋu3 su2 ɲiɪʔ5 pã2 tɔ2 ɣəʔ5 ɲiɲ3 le3 zã3 he2]	The person I bumped into yesterday is coming to Shanghai.
162	我想的男人来上海	[ŋu3 ɕiã2 ɣəʔ5 nø3 ɲiɲ3 le3 zã3 he2]	The man I miss is coming to Shanghai.
163	我认得的男人来上海	[ŋu3 ɲiɲ3 təʔ4 ɣəʔ5 nø3 ɲiɲ3 le3 zã3 he2]	The man I know is coming to Shanghai.
164	我老欢喜的男人来上海	[ŋu3 lɔ3 huø1 ɕi2 ɣəʔ5 nø3 ɲiɲ3 le3 zã3 he2]	The man that I really like is coming to Shanghai.
165	我昨日碰到的男人来上海	[ŋu3 su2 ɲiɪʔ5 pã2 tɔ2 ɣəʔ5 nø3 ɲiɲ3 le3 zã3 he2]	The man that I bumped into yesterday is coming to Shanghai.
166	我骂伊	[ŋu3 mo3 ɦi3]	I scold him.
167	我骂小李	[ŋu3 mo3 ɕiɔ2 li3]	I scold Little Li.
168	伊骂董事长	[ɦi3 mo3 toŋ2 zz3 tsã2]	He scolds the trustee.
169	伊骂女朋友	[ɦi mo3 ɲy3 bəŋ3 ɦiɤ3]	He scolds the girlfriend.
170	伊骂苏州校长	[ɦi3 mo3 su1 tsɤ1 ɕiɔ2 tsã2]	He scolds the Suzhou principal.
171	小余骂伊	[ɕiɔ2 ɦy3 mo3 ɦi3]	Little Yu scolds him.
172	小余骂小李	[ɕiɔ2 ɦy3 mo3 ɕiɔ2 li3]	Little Yu scolds Little Li.
173	小余骂董事长	[ɕiɔ2 ɦy3 mo3 toŋ2 zz3 tsã2]	Little Yu scolds the trustee.
174	小余骂女朋友	[ɕiɔ2 ɦy3 mo3 ɲy3 bəŋ3 ɦiɤ3]	Little Yu scolds the girlfriend.
175	小余骂苏州校长	[ɕiɔ2 ɦy3 mo3 su1 tsɤ1 ɕiɔ2 tsã2]	Little Yu scolds the Suzhou principal.
176	小吴骂搿眼人	[ɕiɔ2 u1 mo3 kəʔ4 ɲe3 ɲiɲ3]	Little Wu scolds these few people.
177	小吴骂五个人	[ɕiɔ2 u1 mo3 ŋ3 ɣəʔ5 ɲiɲ3]	Little Wu scolds five people.
178	小吴骂搿五个人	[ɕiɔ2 u1 mo3 kəʔ4 ŋ3 ɣəʔ5 ɲiɲ3]	Little Wu scolds these five people.
179	小吴骂十五个人	[ɕiɔ2 u1 mo3 zəʔ5 ŋ3 ɣəʔ5 ɲiɲ3]	Little Wu scolds fifteen people.
180	小吴骂搿十五个人	[ɕiɔ2 u1 mo3 kəʔ4 zəʔ5 ŋ3 ɣəʔ5 ɲiɲ3]	Little Wu scolds these fifteen people.
181	小吴骂搿眼小人	[ɕiɔ2 u1 mo3 kəʔ4 ɲe3 ɕiɔ2 ɲiɲ3]	Little Wu scolds these few kids.
182	小吴骂五个小人	[ɕiɔ2 u1 mo3 ŋ3 ɣəʔ5 ɕiɔ2 ɲiɲ3]	Little Wu scolds five kids.
183	小吴骂搿五个小人	[ɕiɔ2 u1 mo3 kəʔ4 ŋ3 ɣəʔ5 ɕiɔ2 ɲiɲ3]	Little Wu scolds these five kids.
184	小吴骂十五个小人	[ɕiɔ2 u1 mo3 zəʔ5 ŋ3 ɣəʔ5 ɕiɔ2 ɲiɲ3]	Little Wu scolds fifteen kids.

	Simplified Characters	IPA	English Translation
185	小吴骂搿十五个小人	[ɕiɔ2 u1 mo3 kəʔ4 zəʔ5 ŋ3 ɣəʔ5 ɕiɔ2 ɲiɲ3]	Little Wu scolds these fifteen kids
186	小李像侬爷	[ɕiɔ2 li3 ʑiã3 na3 ɦia3]	Little Li resembles y'all's dad.
187	小李像阿拉爷	[ɕiɔ2 li3 ʑiã a2 la3 ɦia3]	Little Li resembles our dad.
188	小李喂我的猫	[ɕiɔ2 li3 ue2 ŋu3 ɣəʔ5 mɔ1]	Little Li feeds my cat.
189	小李喂妹妹的猫	[ɕiɔ2 li3 ue2 me3 me3 ɣəʔ5 mɔ1]	Little Li feeds the little sister's cat.
190	小李喂女朋友的猫	[ɕiɔ2 li3 ue2 ɲy3 bəŋ3 ɦiɤ3 ɣəʔ5 mɔ1]	Little Li feeds the girlfriend's cat.
191	小李喂董事长的猫	[ɕiɔ2 li3 ue2 toŋ2 zz3 tsã2 ɣəʔ5 mɔ1]	Little Li feeds the trustee's cat.
192	妹妹像侬姆妈	[me3 me3 ʑiã3 na3 m1 ma1]	Little sister resembles y'all's mom.
193	妹妹像阿拉姆妈	[me3 me3 ʑiã a2 la3 m1 ma1]	Little sister resembles our mom.
194	妹妹像我的爱人	[me3 me3 ʑiã ŋu3 ɣəʔ5 e2 ɲiɲ3]	Little sister resembles my spouse.
195	妹妹像小李的爱人	[me3 me3 ʑiã ɕiɔ2 li3 ɣəʔ5 e2 ɲiɲ3	Little sister resembles Little Li's spouse.
196	妹妹像女朋友的朋友	[me3 me3 ʑiã ɲy3 bəŋ3 ɦiɤ3 ɣəʔ5 bəŋ3 ɦiɤ3]	Little sister resembles the girlfriend's friend.
197	妹妹像董事长的爱人	[me3 me3 ʑiã toŋ2 zz3 tsã2 ɣəʔ5 e2 ɲiɲ3]	Little sister resembles the trustee's spouse.
198	小李喂侬爷的猫	[ɕiɔ2 li3 ue2 na3 ɦia3 ɣəʔ5 mɔ1]	Little Li fees y'all's dad's cat.
199	小李看我的猫的脚	[ɕiɔ2 li3 kʰø ŋu3 ɣəʔ5 mɔ1 ɣəʔ5 tɕiaʔ4]	Little Li sees my cat's foot.
200	小李喂伊妹妹的猫	[ɕiɔ2 li3 ue2 ɦi3 me3 me3 ɣəʔ5 mɔ1]	Little Li feeds his little sister's cat.
201	小李喂伊的爱人的猫	[ɕiɔ2 li3 ue2 ɦi3 ɣəʔ5 e2 ɲiɲ3 ɣəʔ5 mɔ1]	Little Li feed his spouses cat.
202	小李喂伊的女朋友的猫	[ɕiɔ2 li3 ue2 ɦi3 ɣəʔ5 ɲy3 bəŋ3 ɦiɤ3 ɣəʔ5 mɔ1]	Little Li feed his girlfriends cat.
203	小李喂伊的董事长的猫	[ɕiɔ2 li3 ue2 ɦi3 ɣəʔ5 toŋ2 zz3 tsã2 ɣəʔ5 mɔ1]	Little Li feeds his trustee's cat.
204	小李喂侬爷的小猫	[ɕiɔ2 li3 ue2 na3 ɦia3 ɣəʔ5 ɕiɔ2 mɔ1]	Little Li feeds y'all's dad's cat.
205	小李看我的猫的脚爪	[ɕiɔ2 li3 kʰø ŋu3 ɣəʔ5 mɔ1 ɣəʔ5 tsiaʔ4 tsɔ2]	Little Li sees my cat's claw.
206	小李喂伊妹妹的小猫	[ɕiɔ2 li3 ue2 ɦi3 me3 me3 ɣəʔ5 ɕiɔ2 mɔ1]	Little Li feeds his little sister's kitten.
207	小李喂伊的爱人的小猫	[ɕiɔ2 li3 ue2 ɦi3 ɣəʔ5 e2 ɲiɲ3 ɣəʔ5 ɕiɔ2 mɔ1]	Little Li feeds his spouse's kitten.
208	小李喂伊的女朋友的小猫	[ɕiɔ2 li3 ue2 ɦi3 ɣəʔ5 ɲy3 bəŋ3 ɦiɤ3 ɣəʔ5 ɕiɔ2 mɔ1]	Little Li feeds his girlfriend's kitten.
209	小李喂伊的董事长的小猫	[ɕiɔ2 li3 ue2 ɦi3 ɣəʔ5 toŋ2 zz3 tsã2 ɣəʔ5 ɕiɔ2 mɔ1]	Little Li feeds his trustee's kitten.

	Simplified Characters	IPA	English Translation
210	小余想淡的菜	[ɕiɔ2 ɦy3 ɕiã2 de3 ɣəʔ5 tsʰe2]	Little Yu likes light food.
211	小余是健康的人	[ɕiɔ2 ɦy3 zz3 dʑi3 kʰã1 ɣəʔ5 ɲiɲ3]	Little Yu is a healthy person.
212	小余喂毛茸茸的猫	[ɕiɔ2 ɦy3 ue2 mɔ3 ɦioŋ3 ɦioŋ3 ɣəʔ5 mɔ1]	Little Yu feeds the fluffy cat.
213	小余想老淡的菜	[ɕiɔ2 ɦy3 ɕiã2 lɔ3 de3 ɣəʔ5 tsʰe2]	Little Yu likes very light food.
214	小余是老健康的人	[ɕiɔ2 ɦy3 zz3 lɔ3 dʑi3 kʰã1 ɣəʔ5 ɲiɲ3]	Little Yu is a very healthy person.
215	小余想淡的小菜	[ɕiɔ2 ɦy3 ɕiã2 de3 ɣəʔ5 ɕiɔ2 tsʰe2]	Little Yu likes light side dishes.
216	小余是健康的小人	[ɕiɔ2 ɦy3 zz3 dʑi3 kʰã1 ɣəʔ5 ɕiɔ2 ɲiɲ3]	Little Yu is a healthy kid.
217	小余喂毛茸茸的小猫	[ɕiɔ2 ɦy3 ue2 mɔ3 ɦioŋ3 ɦioŋ3 ɣəʔ5 ɕiɔ2 mɔ1]	Little Yu feeds the fluffy kitten.
218	小余想老淡的小菜	[ɕiɔ2 ɦy3 ɕiã2 lɔ3 de3 ɣəʔ5 ɕiɔ2 tsʰe2]	Little Yu likes very light side dishes.
219	小余是老健康的小人	[ɕiɔ2 ɦy3 zz3 lɔ3 dʑi3 kʰã1 ɣəʔ5 ɕiɔ2 ɲiɲ3]	Little Yu is a very healthy kid.
220	小吴骂我打的人	[ɕiɔ2 u1 mo3 ŋu3 tã2 ɣəʔ5 ɲiɲ3]	Little Wu scolds the person I hit.
221	小吴骂我认得的人	[ɕiɔ2 u1 mo3 ŋu3 ɲiɲ3 təʔ4 ɣəʔ5 ɲiɲ2]	Little Wu scolds the person I know.
222	小吴骂我老欢喜的人	[ɕiɔ2 u1 mo3 ŋu3 lɔ3 huø1 ɕi2 ɣəʔ5 ɲiɲ3]	Little Wu scolds the person I like.
223	小吴骂我刚刚碰到的人	[ɕiɔ2 u1 mo3 ŋu3 kã1 kã1 pã2 tɔ2 ɣəʔ5 ɲiɲ3]	Little Wu scolds the person I just bumped into.
224	小吴骂我打的小人	[ɕiɔ2 u1 mo3 ŋu3 tã2 ɣəʔ5 ɕiɔ2 ɲiɲ3]	Little Wu scolds the kid I hit.
225	小吴骂我认得的小人	[ɕiɔ2 u1 mo3 ŋu3 ɲiɲ3 təʔ4 ɣəʔ5 ɕiɔ2 ɲiɲ2	Little Wu scolds the kid I know.
226	小吴骂我老欢喜的小人	[ɕiɔ2 u1 mo3 ŋu3 lɔ3 huø1 ɕi2 ɣəʔ5 ɕiɔ2 ɲiɲ3	Little Wu scolds the kid I really like.
227	小吴骂我刚刚碰到的小人	[ɕiɔ2 u1 mo3 ŋu3 kã1 kã1 pã2 tɔ2 ɣəʔ5 ɕiɔ2 ɲiɲ3]	Little Wu scolds the kid I just bumped into.
228	侬会骂伊	[noŋ3 ɦue3 mo3 ɦi3]	You will scold him.
229	小余会骂伊	[ɕiɔ2 ɦy3 ɦue3 mo3 ɦi3]	Little Yu will scold him.
230	侬会讲英语	[noŋ3 ɦue3 kã2 iɲ1 ɲy3]	You will speak English.
231	小余会讲英语	[ɕiɔ2 ɦy3 ɦue3 kã2 iɲ1 ɲy3]	Little Yu will speak English.
232	侬应该骂伊	[noŋ3 iɲ1 ke1 mo3 ɦi3]	You should scold him.
233	小余应该骂伊	[ɕiɔ2 ɦy3 iɲ1 ke1 mo3 ɦi3]	Little Yu should scold him.
234	侬应该讲英语	[noŋ3 iɲ1 ke1 kã2 iɲ1 ɲy3]	You should speak English.
235	小余应该讲英语	[ɕiɔ2 ɦy3 iɲ1 ke1 kã2 iɲ1 ɲy3]	Little Yu should speak English.

	Simplified Characters	IPA	English Translation
236	小吴勿像小李	[ɕiɔ2 u1 vəʔ5 ʑiã3 ɕiɔ2 li3]	Little Wu doesn’t resemble Little Li.
237	小吴老像小李	[ɕiɔ2 u1 lɔ3 ʑiã3 ɕiɔ2 li3]	Little Wu resembles Little Li a lot.
238	伊啦侪像小李	[ɕiɔ2 u1 lɔ3 ʑiã3 ɕiɔ2 li3]	They all resemble Little Li.
239	小吴明朝来上海	[ɕiɔ2 u1 miɲ3 tsɔ2 le3 zã3 he2]	Little Wu is coming to Shanghai tomorrow.
240	小吴礼拜五来上海	[ɕiɔ2 u1 li3 pa2 ŋ3 le3 zã3 he2]	Little Wu is coming to Shanghai on Friday.
241	小吴明朝夜里厢来上海	[ɕiɔ2 u1 miɲ3 tsɔ2 ɦia3 li3 ɕiã2 le3 zã3 he2]	Little Wu is coming to Shanghai tomorrow night.
242	伊帮我买油条	[ɦi3 pã1 ŋu3 ma3 ɦiɤ3 diɔ3]	He helps me buy a donut.
243	小余帮我买油条	[ɕiɔ2 ɦy3 pã1 ŋu3 ma3 ɦiɤ3 diɔ3]	Little Yu helps me buy a donut.
244	伊嘞上海卖地皮	[ɦi3 ləʔ5 zã3 he2 ma3 di3 bi3]	He sells real estate in$_1$ Shanghai.
245	小余嘞上海卖地皮	[ɕiɔ2 ɦy3 ləʔ5 zã3 he2 ma3 di3 bi3]	Little Yu sells real estate in$_1$ Shanghai.
246	伊嘞饭店里买鳗鱼	[ɦi3 ləʔ5 ve3 ti2 li3 ma3 mø3 n3]	He ordered eel in$_1$ the restaurant.
247	小余嘞饭店里买鳗鱼	[ɕiɔ2 ɦy3 ləʔ5 ve3 ti2 li3 ma3 mø3 n3]	Little Yu ordered eel in$_1$ the restaurant.
248	伊嘞嗨上海卖地皮	[ɦi3 ləʔ5 he2 zã3 he2 ma3 di3 bi3]	He sells real estate in$_2$ Shanghai.
249	小余嘞嗨上海卖地皮	[ɕiɔ2 ɦy3 ləʔ5 he2 zã3 he2 ma3 di3 bi3]	Little Yu sells real estate in$_2$ Shanghai.
250	小余嘞嗨饭店里买鳗鱼	[ɦi3 ləʔ5 he2 ve3 ti2 li3 ma3 mø3 n3]	He ordered eel in$_2$ the restaurant.
251	小余嘞嗨饭店里买鳗鱼	[ɕiɔ2 ɦy3 ləʔ5 he2 ve3 ti2 li3 ma3 mø3 n3]	Little Yu ordered eel in$_2$ the restaurant.
252	小李唱歌唱得老好	[ɕiɔ2 li3 tsʰã2 ku1 tsʰã təʔ4 lɔ3 hɔ2]	Little Li sings well.
253	小李唱歌唱得老好听	[ɕiɔ2 li3 tsʰã2 ku1 tsʰã təʔ4 lɔ3 hɔ2 tʰiɲ1]	Little Li sings beautifully.
254	小李唱歌唱得交关好听	[ɕiɔ2 li3 tsʰã2 ku1 tsʰã təʔ4 tɕiɔ1 kue2 hɔ2 tʰiɲ1]	Little Li sings incredibly beautifully.
255	小李讲英语讲得老好	[ɕiɔ2 li3 kã2 iɲ1 ɲy3 kã təʔ4 lɔ3 hɔ2]	Little Li speaks English well.
256	小李讲英语讲得老清爽	[ɕiɔ2 li3 kã2 iɲ1 ɲy3 kã təʔ4 lɔ3 tɕʰiɲ1 sã2]	Little Li speaks English clearly.
257	小李讲英语讲得交关清爽	[ɕiɔ2 li3 kã2 iɲ1 ɲy3 kã təʔ4 tɕiɔ1 kue1 tɕʰiɲ1 sã2]	Little Li speaks English incredibly clearly.
258	我去看侬	[ŋu3 tɕʰi2 kʰø2 noŋ3]	I go and see you.
259	小李去看侬	[ɕiɔ2 li3 tɕʰi2 kʰø2 noŋ3]	Little Li goes and sees you.
260	小李去看小吴	[ɕiɔ2 li3 tɕʰi2 kʰø2 ɕiɔ2 u1]	Little Li goes and sees Little Wu.
261	我去看小吴	[ŋu3 tɕʰi2 kʰø2 ɕiɔ2 u1]	I go and see Little Wu.

	Simplified Characters	IPA	English Translation
262	洪水退下去了	[ɦoŋ3 sz2 tʰe2 ɦo3 tɕʰi2 ləʔ5]	The flood is retreating.
263	小吴哭起来老快额	[ɕiɔ2 u1 kʰɔʔ4 tɕʰi2 le3 lɔ3 kʰua2 ɣəʔ5]	Little Wu started crying very quickly.
264	搿句读起来勿清爽额	[kəʔ4 tɕy2 dɔʔ5 tɕʰi2 le3 vəʔ5 tɕʰiɲ1 sã2 ɣəʔ5]	This sentence wasn't read clearly.
265	小吴拿我打下去了	[ɕiɔ2 u1 na3 ŋu3 tã2 ɕia2 tɕʰi2 ləʔ5]	Little Wu beat me down.
266	小吴拿文章写出来了	[ɕiɔ2 u1 na3 vəŋ3 tsã1 ɕia2 tsʰəʔ4 le3 ləʔ5]	Little Wu finished writing the essay.
267	小吴英语讲得来老快额	[ɕiɔ2 u1 iɲ1 ɲy3 kã2 le3 lɔ3 kʰua2 ɣəʔ5]	Little Wu speaks English very fast.
268	小吴搿个字讲出来勿清爽	[ɕiɔ2 u1 kəʔ4 kəʔ4 tsz2 kã2 tsʰəʔ4 le3 vəʔ5 tɕʰiɲ1 sã2]	Little Wu didn't read this character clearly.
269	小李晓得伊骂人	[ɕiɔ2 li3 ɕiɔ2 tə24 ɦi3 mo3 ɲiɲ3]	Little Li knows he scolds people.
270	小李晓得小余骂人	[ɕiɔ2 li3 ɕiɔ2 tə24 ɕiɔ2 ɦy3 mo3 ɲiɲ3]	Little Li knows Little Yu scolds people.
271	小李晓得伊卖地皮	[ɕiɔ2 li3 ɕiɔ2 tə24 ɦi3 ma3 di3 bi3]	Little Li knows he sells real estate.
272	小李晓得小余卖地皮	[ɕiɔ2 li3 ɕiɔ2 tə24 ɕiɔ2 ɦy3 ma3 di3 bi3]	Little Li knows Little Yu sells real estate.
273	小李晓得伊有女朋友	[ɕiɔ2 li3 ɕiɔ2 tə24 ɦi3 ɦiɤ3 ɲy3 bəŋ3 ɦiɤ3]	Little Li knows he has a girlfriend.
274	小李晓得小余有女朋友	[ɕiɔ2 li3 ɕiɔ2 tə24 ɕiɔ2 ɦy3 ɦiɤ3 ɲy3 bəŋ3 ɦiɤ3]	Little Li knows Little Yu has a girlfriend.
275	假使伊来，阿拉就走	[ka1 zz3 ɦi3 | le3 a2 la2 dʑiɤ3 tsɤ2]	If he comes, we leave.
276	假使小吴来，阿拉就走	[ka1 zz3 ɕiɔ2 u1 le3 | a2 la2 dʑiɤ3 tsɤ2]	If Little Wu comes, we leave.
277	假使伊来，小李就要走	[ka1 zz3 ɦi3 le3 | ɕiɔ2 li3 dʑiɤ3 iɔ2 tsɤ2]	If he comes, Little Li will just want to leave.
278	假使小吴来，小李就要走	[ka1 zz3 ɕiɔ2 u1 le3 | ɕiɔ2 li3 dʑiɤ3 iɔ2 tsɤ2]	If Little Wu comes, Little Li will just want to leave.
279	假使侬有问题，侬可以叫我	[ka1 zz3 noŋ3 ɦiɤ3 vəŋ3 ti2 | noŋ3 kʰu2 i2 tɕiɔ2 ŋu3]	If you have problems, you can call me.
280	假使有辰光，应该去看苏州	[ka1 zz3 ɦiɤ3 zəŋ3 kuã1 | iɲ1 ke1 tɕʰi2 kʰø2 su1 tsɤ2]	If [you] have time, [you] should go see Suzhou.
281	假使侬有辰光，应该去看苏州	ka1 zz3 noŋ3 ɦiɤ3 zəŋ3 kuã1 iɲ1 ke1 tɕʰi2 kʰø2 su1 tsɤ2]	If you have time, [you] should go see Suzhou.
282	假使小李有辰光，跟伊讲讲闲话	[ka1 zz3 ɕiɔ2 li3 ɦiɤ3 zəŋ3 kuã1 kən1 ɦi3 kã2 kã2 ɦe3 ɦu3]	If Little Li has time, chat with him.

283	假使侬有辰光，应该跟伊讲闲话	[ka1 zz3 noŋ3 ɦiɤ3 zən3 kuã1 iɲ1 ke1 kən1 ɦi3 kã2 ɦe3 ɦu3]	If you have time, [you] should chat with him.

References and Consulted Works

Arnhold, A. (2014). Prosodic structure and focus realization in West Greenlandic. In Sun-Ah Jun (ed.) *Prosodic Typology II: The Phonology of Intonation and Phrasing.* Oxford University Press.

Beckman, M. E., Hirschberg, J., & Shattuck-Hufnagel, S. (2005). The original ToBI system and the evolution of the ToBI framework. In Sun-Ah Jun (ed.) *Prosodic Typology–The Phonology of Intonation and Phrasing.* Oxford, OUP.

Beckman, Mary E., & Hirschberg, J. (1994). The ToBI annotation conventions. *Ohio State University*.

Beckman, Mary E., & Pierrehumbert, J. B. (1986). Intonational structure in Japanese and English. *Phonology*, *3*, 255–309.

Boersma, P. (1997). How we learn variation, optionality, and probability. *Proceedings of the Institute of Phonetic Sciences of the University of Amsterdam*, *21*, 43–58.

Boersma, P., & Weenink, D. (2013). Praat: Doing phonetics by computer [Computer program]. Version 5.3. 51. *Online: Http://Www. Praat. Org/Retrieved, Last Viewed On*, *12*.

Bruce, G. (1977). *Swedish word accents in sentence perspective* (Vol. 12). LiberLäromedel/Gleerup Malmö.

Bruce, G. (1982). Developing the Swedish intonation model. *Journal ISSN*, *280*, 526X.

Bruce, G. (2005). Intonational prominence in varieties of Swedish revisited. In Sun-Ah Jun (ed.) *Prosodic Typology–The Phonology of Intonation and Phrasing.* Oxford, OUP. 410–429.

Cao J. F., & Maddieson, I. (1992). An exploration of phonation types in Wu dialects of Chinese. *Journal of Phonetics*, *20*(1), 77–92.

Chahal, D., & Hellmuth, S. (2014). *The Intonation of Lebanese and Egyptian Arabic.* In Sun-Ah Jun (ed.) Prosodic Typology II: The Phonology of Intonation and Phrasing. Oxford University Press. Pp. 365-404.

Chan, M. K. (1995). An autosegmental analysis of Danyang tone sandhi: Some historical and theoretical issues. *Wuyu Yanjiu [Studies of the Wu Dialects]*, 145–184.

Chao, Y. R. (1928). Xiandai Wuyu Yanju [Studies in the Modern Wu dialects]. *Peking: Tsing Hua College Research Institute, Monograph*, *4*.

Chao, Y. R. (1948). *Mandarin Primer* Cambridge*. Mass.: Harvard University Press.

Chen, M. Y. (2000). *Tone sandhi: Patterns across Chinese dialects* (Vol. 92). Cambridge University Press.

Chen, Y. (2008). Revisiting the phonetics and phonology of Shanghai tone sandhi. *Proceedings of Speech Prosody 2008*, 253–256.

Chen, Y. (2009). Prosody and information structure mapping: Evidence from Shanghai Chinese. *Chinese Journal of Phonetics*, *2*(1), 123–133.

Chen, Y. (2011). How does phonology guide phonetics in segment–f0 interaction? *Journal of Phonetics*, *39*(4), 612–625.

Chen, Y., & Gussenhoven, C. (2015). Shanghai Chinese. *Journal of the International Phonetic Association*, *45*(3), 321–337.

Chen, Z. (2010). An acoustic study of voiceless onset followed by breathiness of Wu dialects: Based on the Shanghai dialect. *Yuyan Yanjiu*, *30*(3), 20–34.

Duanmu, S. (1993). Rime length, stress, and association domains. *Journal of East Asian Linguistics*, *2*(1), 1–44.

Duanmu, S. (1995). Metrical and tonal phonology of compounds in two Chinese dialects. *Language*, 225–259.

Duanmu, S. (1997). Recursive constraint evaluation in Optimality Theory: Evidence from cyclic compounds in Shanghai. *Natural Language & Linguistic Theory*, *15*(3), 465–507.

Duanmu, S. (1999). Metrical structure and tone: Evidence from Mandarin and Shanghai. *Journal of East Asian Linguistics*, *8*(1), 1–38.

El Zarka, D. (2013). *On the interaction of information structure and prosody: The case of Egyptian Arabic*.

Elordieta, G. (1997). Accent, tone and intonation in Lekeitio Basque. *Issues in the Phonology and Morphology of the Major Iberian Languages*, *378*.

Esposito, C. M. (2012). An acoustic and electroglottographic study of White Hmong tone and phonation. *Journal of Phonetics*, *40*(3), 466–476.

Estebas-Vilaplana, E., & Prieto, P. (2010). Castilian Spanish intonation. *Transcription of Intonation of the Spanish Language*, *17*, 48.

Gao, J. (2016). Sociolinguistic motivations in sound change: On-going loss of low tone breathy voice in Shanghai Chinese. *Papers in Historical Phonology*, *1*, 166–186.

Gao, J., & Hallé, P. (2016). Prosodic organization and microprosodic effects in Shanghai Chinese. *15th Laboratory Phonology*.

Gao, J., & Hallé, P. (2013a). Are young male speakers losing Tone 3 breathiness in Shanghai Chinese? An acoustic and electroglottographic study. *Proceedings of the 2nd International Conference on Phonetics of the Languages in China*, 163–166.

Gao, J., & Hallé, P. A. (2013b). Duration as a secondary cue for perception of voicing and tone in shanghai Chinese. *INTERSPEECH*, 3157–3161.

Gao, J., & Hallé, P. A. (2015). The role of voice quality in Shanghai tone perception. *ICPhS*.

Gao J. Y., & Hallé, P. (2017). Phonetic and phonological properties of tones in Shanghai Chinese. *Cahiers de Linguistique Asie Orientale*, *46*(1), 1–31.

Gao, J., Hallé, P., & Draxler, C. (2020). Breathy voice and low-register: A case of trading relation in Shanghai Chinese tone perception? *Language and Speech*, *63*(3), 582–607.

Gao, J., Hallé, P., Honda, K., Maeda, S., & Toda, M. (2011). *Shanghai slack voice: Acoustic and EPGG data*.

Garellek, M., & Keating, P. (2011). The acoustic consequences of phonation and tone interactions in Jalapa Mazatec. *Journal of the International Phonetic Association*, 185–205.

Godjevac, S. (2005). Transcribing Serbo-Croatian intonation. In Sun-Ah Jun (ed.) *Prosodic Typology–The Phonology of Intonation and Phrasing.* Oxford, OUP. 146–171.

Goldsmith, J. (1976). *Autosegmental phonology*. MIT Press London.

Goldwater, S., & Johnson, M. (2003). Learning OT constraint rankings using a maximum entropy model. *Proceedings of the Stockholm Workshop on Variation within Optimality Theory*, *111120*.

Grice, Martine, S. Baumann & R. Benzmüller (2005). German Intonation in Autosegmental-metrical Phonology. In Jun, Sun-Ah (Eds.*). Prosodic Typology: The Phonology of Intonation and Phrasing.* Oxford: OUP. 55-83.

Haudricourt, A. G. (1954). Comment reconstruire le chinois archaïque. *Word*, *10*(2–3), 351–364.

Haudricourt, A. G. (1961). Richesse en phonèmes et richesse en locuteurs. *L'homme*, 5–10.

Hayes, B. (1988). Metrics and phonological theory. *Linguistics: The Cambridge Survey*, *2*, 220–249.

Hellmuth, S. J. (2006). *Intonational pitch accent distribution in Egyptian Arabic.* SOAS University of London.

Hillenbrand, J., Cleveland, R. A., & Erickson, R. L. (1994). Acoustic correlates of breathy vocal quality. *Journal of Speech, Language, and Hearing Research*, *37*(4), 769–778.

Huang, C. T. J. (1982). Move wh in a language without wh-movement. *The linguistic review*, 1(4), 369-416.

Hyman, L. M. (2006). Word-prosodic typology. *Phonology*, 225–257.

Hyman, L. M. (2007). Tone: is it Different?. *UC Berkeley Phonlab Annual Report, Vol. 3.* 483-528.

Jin, S. (1986). *Shanghai morphotonemics: A preliminary study of tone sandhi behavior across word boundaries* (Vol. 327). Indiana University Linguistics Club.

Jongman, A., Wang, Y., Moore, C. B., & Sereno, J. A. (2006). *Perception and production of Mandarin Chinese tones*. Submitted to: Handbook of Chinese Psycholinguistics. E. Bates, L.H. Tan, and O.J.L Tzeng (eds.). Cambridge University Press.

Jun, S.-A. (1993). The phonetics and phonology of Korean prosody (Ph. D. dissertation). *New York: Ohio State University.(Published 1996 by Garland Publishing Inc.).*

Jun, S.-A. (1996). *The phonetics and phonology of Korean prosody: Intonational phonology and prosodic structure*. Taylor & Francis.

Jun, S.-A. (1998). The accentual phrase in the Korean prosodic hierarchy. *Phonology*, 189–226.

Jun, S.-A. (2000). K-ToBI (Korean ToBI) labelling conventions. *Speech Sciences*, *7*(1), 143–170.

Jun, S.-A. (2005). Korean intonational phonology and prosodic transcription. In Sun-Ah Jun (ed.) *Prosodic Typology–The Phonology of Intonation and Phrasing.* Oxford, OUP. 201.

Jun, S.-A. (2011) Prosodic Markings of Complex NP Focus, Syntax, and the Pre-/Post-Focus String. In Proceedings of the 28th WCCFL, pp. 214-230.Jun, S. A., & Fougeron, C. (2002). Realizations of accentual phrase in French intonation. *Probus*, *14*(1), 147–172.

Karlgren, B. (1922). The reconstruction of ancient Chinese. *T'oung Pao*, *21*(1), 1–42.

Khan, S. ud D. (2012). The phonetics of contrastive phonation in Gujarati. *Journal of Phonetics*, *40*(6), 780–795.

Krom, G. de. (1993). A cepstrum-based technique for determining a harmonics-to-noise ratio in speech signals. *Journal of Speech, Language, and Hearing Research*, *36*(2), 254–266.

Kuang, J., & Keating, P. (2014). Vocal fold vibratory patterns in tense versus lax phonation contrasts. *The Journal of the Acoustical Society of America*, *136*(5), 2784–2797.

Kuang, J., & Tian, J. (2019). *TONE REPRESENTATION AND TONE PROCESSING IN SHANGHAINESE*.

Kuang, J., Tian, J., & Zhou, Y. (2018). The common word prosody in Northern Wu. *Proceedings of the 6th International Symposium on Tonal Aspects of Language*, 7–11.

Ladd, D. R. (2008). *Intonational phonology*. Cambridge University Press.

Ladefoged, P., & Maddieson, I. (1996). *The sounds of the world's languages* (Vol. 1012). Blackwell Oxford.

Li, J., Hu, G., & Wang, R. (2004). Prosody phrase break prediction based on maximum entropy model. *Journal of Chinese Information Processing*, *5*.

Liberman, M., & Prince, A. (1977). On stress and linguistic rhythm. *Linguistic Inquiry*, *8*(2), 249–336.

Liberman, M. Y. (1975). *The intonational system of English.* Dissertation. Massachusetts Institute of Technology.

Ling, B., & Liang, J. (2015). Focus and tonal implementation in Shanghai Chinese. *Chinese Accents and Accented Chinese CAAC2014*, 56.

Ling, B., & Liang, J. (2016). Organizing Syllables into Sandhi Domains-Evidence from F0 and Duration Patterns in Shanghai Chinese. *INTERSPEECH*, 72–76.

Liu, F. (1924). Szu Sheng Shih Yen Lu. Ch'un Yi Shanghai.

Moore, C. B. (1993). Phonetic observations on stress and tones in Mandarin Chinese. *Working Papers of the Cornell Phonetics Laboratory*, *8*, 87–117.

Nespor, M., & Vogel, I. (1986). Prosodic Phonology, Foris, Dordrecht. *NesporProsodic Phonology1986*.

Nespor, M., & Vogel, I. (2007). *Prosodic phonology: With a new foreword* (Vol. 28). Walter de Gruyter.

Peng, S., Chan, M. K., Tseng, C., Huang, T., Lee, O. J., & Beckman, M. E. (2005). Towards a Pan-Mandarin system for prosodic transcription. In Sun-Ah Jun (ed.) *Prosodic Typology–The Phonology of Intonation and Phrasing.* Oxford, OUP. 230–270.

Pierrehumbert, J. B. (1980). *The phonology and phonetics of English intonation*. Massachusetts Institute of Technology.

Pierrehumbert, J., & Beckman, M. (1988). Japanese tone structure. *Linguistic Inquiry Monographs*, *15*, 1–282.

Pulleyblank, E. G. (1978). The nature of the Middle Chinese tones and their development to Early Mandarin. *Journal of Chinese Linguistics*, *6*(2), 173–203.

Qian, N. (2003). Shanghai Yuyan Fazhanshi [A History of the Development of the Shanghai Dialect]. *Shanghai: Shanghai People's Publishing House*.

Qian, N. R., Xu B. H., & Tang Z. Z. (2007). *Shanghaihua da cidian* (Big Shanghainese Dictionary). Shanghai: Shanghai Cishu Chubanshe.

Ramsey, S. R. (1989). *The languages of China*. Princeton University Press.

Ren, N. (1993). *Phonation types and stop consonant distinctions: Shanghai Chinese.*

Ren, N., & Mattingly, I. G. (1989). Spectral slope as a cue for the perception of breathy and non-breathy stops in Shanghainese. *The Journal of the Acoustical Society of America*, *86*(S1), S102–S102.

Roberts, B. (2016). *A Preliminary Model of Shanghainese Intonational Phonology*. MA thesis. University of California, Los Angeles.

Rose, P. (1990). Acoustics and phonology of complex tone sandhi. *Phonetica*, *47*(1–2), 1–35.

Ruan, H. H. (2010). *Xinban zixue shanghaihua* (New Edition Teach Yourself Shanghainese). Shanghai: Shanghai Daxue Chubanshe,

Selkirk, E. (1984). *Phonology and syntax: The relation between sound and structure*. Cambridge, Mass: MIT Press.

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

Selkirk, E., & Shen, T. (1990). *Prosodic domains in Shanghai Chinese. The Phonology-Syntax Connection, edited by Sharon Inkelas & Draga Zec, 313-337*. Chicago: University of Chicago Press.

Shen, T. (1981). Lao pai Shanghai fang yan de lian du bian diao.[Tone sandhi in Old Shanghai.]. *Fangyan*, *81*, 131–144.

Shen, X. S. (1993). Relative duration as a perceptual cue to stress in Mandarin. *Language and Speech*, *36*(4), 415–433.

Sherard, M. (1972). *Shanghai Phonology. Cornell U*. dissertation.

Silverman, K., Beckman, M., Pitrelli, J., Ostendorf, M., Wightman, C., Price, P., Pierrehumbert, J., & Hirschberg, J. (1992). ToBI: A standard for labeling English prosody. *Second International Conference on Spoken Language Processing*.

Sun, L., & Chen, Y. (2015). Post-focus pitch register lowering as a phrasal marker.

‘t Hart, J., & Collier, R. (1975). Integrating Different Levels of Intonation Analysis. *Journal of Phonetics*, *3*(4), 235–255.

Takahashi, Y. (2013a). *The phonological structure of Shanghai tone sandhi*. Ph. D. dissertation, Tokyo University of Foreign Studies, Tokyo, Japan.

Takahashi, Y. (2015). Morpho-Syntactic Effects on Tone Sandhi Variants in Shanghai Chinese. *言語研究 (Gengo Kenkyu)*, *147*, 57–70.

Takahashi, Y. (2013b). The phonetic change of tone sandhi in Shanghai Chinese. *Proceedings of the International Conference on Phonetics of the Languages in China*, 139–142.

Tian, J., & Kuang, J. (2016). Revisiting the register contrast in Shanghai Chinese. *Tonal Aspects of Languages 2016*, 147–151.

Tian, J., & Kuang, J. (2017). Phonetic properties of the non-modal phonation in Shanghainese register contrast. *The Journal of the Acoustical Society of America*, *141*(5), 3982–3983.

Tian, J., & Kuang, J. (2019). The phonetic properties of the non-modal phonation in Shanghainese. *Journal of the International Phonetic Association*, 1–27.

Toda, T. (1990). Shanghai tonal phonology “rightward spreading?”. *Proceedings of the Third Australian International Conference on Speech Science and Technology. Melbourne*.

Venditti, J. J. (2005). The J_ToBI model of Japanese intonation. *Prosodic Typology: The Phonology of Intonation and Phrasing*, *172*, 200.

Wong, W. Y. P., Chan, M. K., & Beckman, M. E. (2005). An autosegmental-metrical analysis and prosodic annotation conventions for Cantonese. *Prosodic Typology: The Phonology of Intonation and Phrasing*, *1*, 271.

Xu, B. H., & Tang, Z. Z. (1988). *Shanghai Shiqu Fangyan Zhi* (A Description of the Urban Shanghai Dialect). Shanghai: Shanghai Educational Publishing House.

Xu, B., Tang, Z., & Qian, N. (1981). Xinpai Shanghai fangyan de liandu biandiao [Tone sandhi in New Shanghai]. *Fangyan (Dialects)*, *1981*, 145–155.

Yan, H., Chien, Y.-F., & Zhang, J. (2020). Priming the representation of left-dominant sandhi words: A Shanghai dialect case study. *Language and Speech*, *63*(2), 362–380.

Yan, H., & Zhang, J. (2016). Pattern substitution in Wuxi tone sandhi and its implication for phonological learning. *International Journal of Chinese Linguistics*, *3*(1), 1–44.

Yip, M. J. (1980). *The tonal phonology of Chinese*. Massachusetts Institute of Technology.

Yoon, T.-J. (2007). *A predictive model of prosody through grammatical interface: A computational approach*. University of Illinois at Urbana-Champaign.

Yue-Hashimoto, A. O. (1987). Tone sandhi across Chinese dialects. *Chinese Language Society of Hong Kong (Ed.), Wang Li Memorial Volumes, English Volume*, 445–474.

Zee, E., & Maddieson, I. (1979). Tones and tone sandhi in Shanghai: Phonetic evidence and phonological analysis. *UCLA Working Papers in Phonetics*, *45*, 93–129.

Zee, E., & Xu L. J. (2016). Shanghainese. In Thurgood, G., & LaPolla, R. J. (Eds.). *The Sino-Tibetan languages.* Taylor & Francis.

Zhang, J. (2007). A directional asymmetry in Chinese tone sandhi systems. *Journal of East Asian Linguistics*, *16*(4), 259–302.

Zhang, J., & Meng, Y. (2016). Structure-dependent tone sandhi in real and nonce disyllables in Shanghai Wu. *Journal of Phonetics*, *54*, 169–201.

Zhang, J., & Yan, H. (2018). Contextually dependent cue realization and cue weighting for a laryngeal contrast in Shanghai Wu. *The Journal of the Acoustical Society of America*, *144*(3), 1293–1308.

Zhu, X. (1995). *Shanghai tonetics*. Lincom Europa München.

Zhu, X. (2006). *A grammar of Shanghai Wu* (Vol. 66). Lincom Europa München.