The Tonal and Intonational Phonology of Lhasa Tibetan

Keh Sheng Lim

A thesis submitted in partial fulfillment of the requirements for the
Doctorate in Philosophy degree in Linguistics
Department of Linguistics

Faculty of Graduate and Postdoctoral Studies
University of Ottawa

Abstract

This dissertation provides a comprehensive description of the tonal and intonational phonology of Lhasa Tibetan (LT) in the Autosegmental-Metrical framework. It is based on recorded data elicited from members of the Tibetan-Canadian community in Ottawa and Toronto.

The first two chapters of the dissertation contain background information about LT, a summary of previous research on LT tones and intonation, and an overview of the theoretical framework and conceptual tools used in the rest of the dissertation.

The third chapter deals with word tonology. I establish that the prosodic structure of LT brings evidence for four main constituents at or below the word level: a) the mora encodes vowel length contrasts, b) the syllable is the tone-bearing unit (TBU), c) the prosodic word, which is maximally binary, delimits the application of most tonal processes, and d) the prosodic word group, which matches grammatical words, is the domain of downstep. This prosodic structure provides evidence against the universality of the Prosodic Hierarchy (Selkirk, 2002; Nespor and Vogel, 2007) in that it has no phonological phrase, but has two word-level constituents. I then argue that LT has three lexical tones (H, LH, and L) – L being limited to some suffixes – and propose that these lexical tones are subject to tone rules applying within the prosodic word and the prosodic word group. These tone rules are similar to those proposed by Duanmu (1992), but have been improved to accurately predict the tone patterns of long polysyllabic words. Based on phonetic evidence, I also come to the conclusion that LT no longer has stress, and that the stress pattern found in other Bodic varieties has been reinterpreted as a part of the tonal system.

The fourth chapter analyses phrasal prosody. I argue that LT forms intonational phrases around clauses and marks them with final lengthening, pitch reset, and a limited set of boundary tones (H% and L%). Although communicative func-

tions and information structure are mostly realized by means of final particles and morphosyntactic devices in LT, I show that boundary tones, focal tones, and deaccenting interact with word tones to form complex melodic patterns.

In the fifth chapter, I present a phonologically-based F0 synthesis model to verify the adequacy of the proposed Autosegmental-Metrical model of LT. This F0 synthesis model consists of three main components: a) the tonal targets defined in previous chapters, b) an F0 interpolation component based on the PENTA model (Xu, 2004), and c) an evaluation component allowing a comparison of the F0 contours of real LT utterances with resynthesized F0 contours of the same utterances. The F0 synthesis model is able to generate F0 contours that approximate the F0 contours of real LT recordings, suggesting that the proposed phonological model adequately captures the overall tonal and intonational phonology of LT.

Acknowledgements

My sincere thanks to Jurme Wangda, my Tibetan friend, teacher, and consultant. With his generous contribution of time, he helped me to connect with all of the Tibetan participants in this research. I am grateful to my Tibetan participants for their kindness and patience during the research, and for helping me to understand their language. I would particularly like to remember one of the participants, who will not have the opportunity to see the results of his contribution. His speech data, along with the data from the others, forms the foundation of this research. I will always be indebted to all of my participants for sharing their language and hospitality with me, and giving me a glimpse of their great culture.

My work would have gone nowhere had it not been for the instruction and guidance of Dr. Marc Brunelle at the University of Ottawa—my supervisor for the last nine years. Marc never hesitated to critique my writing when it was not clear, and helped me to keep my thoughts organized. If what I have written is comprehensible, it is due to his nuanced insightful comments on each and every draft. It was a privilege to have Marc as my supervisor. His relentless commitment to his students is exemplary and I am very grateful to him. I would like to thank Dr. Andrés Salanova, who provided valuable feedback on an earlier draft of this dissertation. Dr. Salanova's encouragement and friendship are deeply appreciated.

I would also like to express my gratitude towards my committee members: Dr. Lev Blumenfeld, Dr. Andrés Salanova, Dr. Kevin McMullin, and Dr. Nancy Caplow. They have provided significant insights and critiques that not only improve the quality of this dissertation but also inspire future research topics. I would like to specifically thank Dr. Nancy Caplow for generously contributing her expertise in verifying the Tibetan data in this dissertation. Her feedback has adverted some significant embarrassments and has helped me improved the readability of transcriptions for other researchers.

Much of my early interest in tonal phenomena came through Dr. Keith Snider, professor of Linguistics at the Canada Institute of Linguistics. He introduced me to tonal phenomena, which have been central to my research interests for the good part of the last decade. He encouraged me to pursue my doctorate, and I am forever grateful to him.

I would also like to thank my editor, Lyra Magloughlin, for her invaluable help in spotting grammatical and stylistic infelicities in my original manuscript.

Even with the support of all these people I could never have completed this dissertation without the support and encouragement of my wife, Mabel Lee. I am forever in debt to you! I am also grateful to our children, Magdalene and Jared, who have shown great understanding and patience. Last but not least, I would like to sincerely thank my parents, Kim Tee and Chian Seng, who have always supported my endeavours.

Contents

List of Figures

List of Tables

Abbreviation

ABL Ablative

ADC Analogue to digital converter

AM Autosegmental Metrical Phonology

ANOM Analysis of Means

ANOVA Analysis of Variance

AVG Average

C Consonant

CT Classical Tibetan

CP Connecting particles

DF Degree of Freedom

DIR Direct

EGO Egocentric

F0 pitch or fundamental frequency

F0R Pitch range

F0z Normalized pitch

FTA F0 target alignment

H High Tone

hon. Honorific

HR Homonym rate

IMP Imperative

I Sound intensity level

IDIR Indirect

Iz Normalized intensity level

L Low Tone

L1 First language

L2 Second language

LH Low High Tone

LSI Lhasa intonation tone rules

LT Lhasa Tibetan

LTF0 LT F0 synthesis model

N Numbers of Sample

O Obstruent

OT Old Tibetan

PO Phonological Output

PWd Prosodic Word

Q interrogative particle

qTA quantitative Target Approximation

R Sonorant

RLS Revised Lhasa Tibetan tone rules

SD Standard Deviation

SF Surface Form

SSQ Sum of Squares

TA Target Approximation model

TAM Tense/Aspect/Modality

UF Underlying Form

V Vowel

WT Written Tibetan or literary Tibetan

Chapter 1

Introduction

Speakers of Tibetan languages can be found in a vast region in the western part of East Asia stretching along both sides of the Himalayan Mountains. Lhasa Tibetan (LT) is an SOV, agglutinative, tonal Central-Tibetan language of the Bodic family. It is the *lingua franca* of the Tibetan people. According to Ethnologue (Lewis et al., 2014), there are 1.172 million LT speakers worldwide, and 920,000 of them are monolingual. These speakers reside mainly in the Tibetan Autonomous Region (TAR) and Xinjiang Province in China. The geographical location of Lhasa is shown in Figure 1.1.

Many LT speakers can be found in various Tibetan diasporic communities in India along the Tibetan border of Himachal Pradesh, Uttarakhand, Arunachal Pradesh, Assam, Delhi, and Sikkim. There is also a significant population of LT speakers in Nepal, residing mainly in Kathmandu, Pokhara, and smaller refugee communities along the Nepal-China border. Many Tibetan diasporas were established in Europe and North America after the Chinese Communist takeover of 1959. In fact, the Canadian diaspora is one of the largest outside of Asia. Since the first Tibetan immigrant landed in Canada in March 1971, at least 7,500 Tibetan and Tibetan Canadians have settled in eleven communities in Ontario, Quebec, Alberta, and

Figure 1.1: The geographical location of Lhasa Tibetan (Simons and Fenning, 2017).

Manitoba (Raska, 2013). A great number of them reside in the 'Tibetan Town' of Parkdale in Toronto. The Canadian Tibetan Cultural Centre was established in Etobicoke, west of Toronto in 2007.

Most mainstream researchers recognize that there are two contrasting tonemes in LT: High and Low. That said, the analysis of the tonal inventory of LT is still controversial, even after a century of investigation[1]. In the following chapters, I claim that some of the previous analyses of LT do not give an accurate representation of LT tonal phonology. While there are many properly described pitch patterns, there is no evidence to support LT developing into a highly complex tone system, as proposed by some authors. In general, this erroneous conclusion is due to a failure to distinguish surface realization or allotones explicitly from tonemes. In this dissertation, I propose a new description of the word tonology and intonational phonology of LT, which incorporates three main research methodologies: *a*) a descriptive phonological analysis, *b*) an instrumental investigation of the F0 pattern, and *c*) a phonetic implementation of the phonological model by means of

[1]The proposed inventories range from atonal to eight tones (Sprigg, 1993; Hu, 2002).

F0 synthesis.

The word tonology and intonational phonology established in this dissertation are based on the instrumental study of the F0 pattern of utterances produced by LT speakers. The dissertation provides a description of the tonal and intonational pitch patterns observed among these speakers. I claim that there are three basic lexical tones: L, LH, and H. H can be found in both stems and affixes, LH can only be found in stems, and L can only be found in affixes. All surface pitch patterns that deviate from these three basic lexical tones are the result of tonal processes affecting lexical tones. Furthermore, I show that it is imperative to take into account the interaction between the lexical tones and intonational tones. Based on the results of my instrumental studies, I conclude that intonational tones modulate the surface realization of lexical tones. The tonal domains of both lexical and intonational tones are governed by LT's prosodic structure. Finally, I present a phonology-based F0 synthesis model to verify the adequacy of the proposed word tonology and intonational phonology.

The remainder of this first chapter provides the reader with essential background information on LT that will facilitate the comprehension of the linguistic discussion in following chapters. §1.1 provides a typological characterization of LT (§1.1), the sociostylistic registers (§1.1.1), a history of LT (§1.1.2), and a synopsis of early research on Tibetan language (§1.1.3). Then, I provide a phonological sketch for LT (§1.2) and a concise discussion of the tonogenesis of LT (§1.3). I conclude this chapter with an outline of the organization of the remainder of the dissertation in §1.4.

1.1 Lhasa Tibetan (LT)

In this section, I provide a relatively concise discussion of the typology of LT, the various registers of LT, the history of the development of the Tibetan language, the traditional view of tonal phenomena, and the early linguistic research conducted by Western scholars on Tibetan.

The Tibetic languages form one of the major language families in the Trans-Himalayan region, to which Classical Literary Tibetan is affiliated. According to Tournadre (2013), there are eight principal subgroups of Tibetic languages. LT are geographically and genetically belongs to the Central subgroup as shown in Table 1.1.

Subgroup	**Languages**
North-Western	Ladakhi, Zanskari, Balti, and Purki.
Western	Spiti, Garzha, Khunu, and Jad.
Central	**Ü-Tsang (LT)**, Phenpo, Lhokha, Tö, and Kongpo
South-Western	Sherpa, Jirel, Humla, Mugu, Dolpo, Lo-ke, Nubri, Tsum, Langtang, Kyirong, Yolmom Gyalsumdo, Kagate, Lhomi, Walung, and Tokpe Gola.
Southern	Dzongkha, Drengjong, Tsamang, Dhromo Lakha, Dur Brokkat, and Mera Sakteng Brokpa-ke
South-Eastern	Hor Nagchu, Hor Bachen, Yushu, Pembar, 'Northern route', Rongdrak, Minyak, 'Southern route', Dzayul, Derong-Jol, Chaktreng, Muli-Dappa, Semkyi Nyida.
Eastern	Drugchu, Khö, Thewo, Chone, Baima, Sharkhok, Palkyi [Pashi] and Zhongu
North Eastern	Amdo, gSerpa, Khalong

Table 1.1: The geolinguistic continuum of Tibetic languages
(Tournadre, 2013, p.122-123)

Based on language family affiliation, LT is a central Tibetan Bodic language of the Western Himalayan group. It is affiliated with the Bodic branch of the Tibeto-Burman language family, as shown in Figure 1.2.

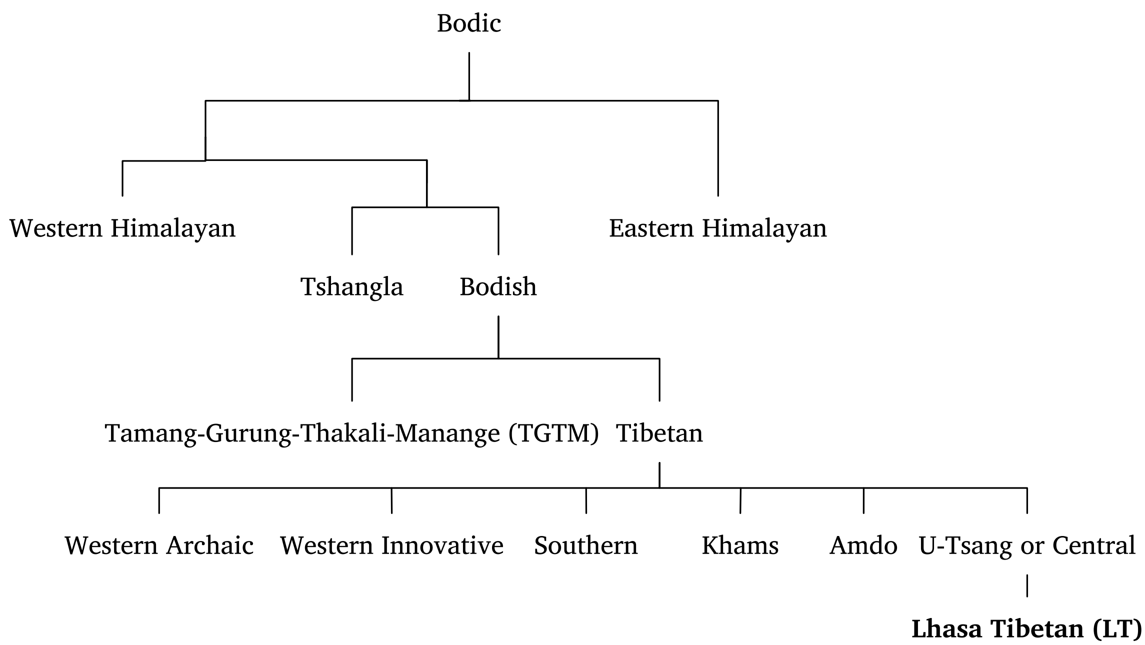


Figure 1.2: The Bodic branch (cf. DeLancey, 2003a,b).

Based on the division between tonal and atonal dialectal groups, LT belongs to the tonal dialectal groups, as shown in Figure 1.3. The classification of Tibetan languages based on tonal phenomena coincides with the genetic classification.

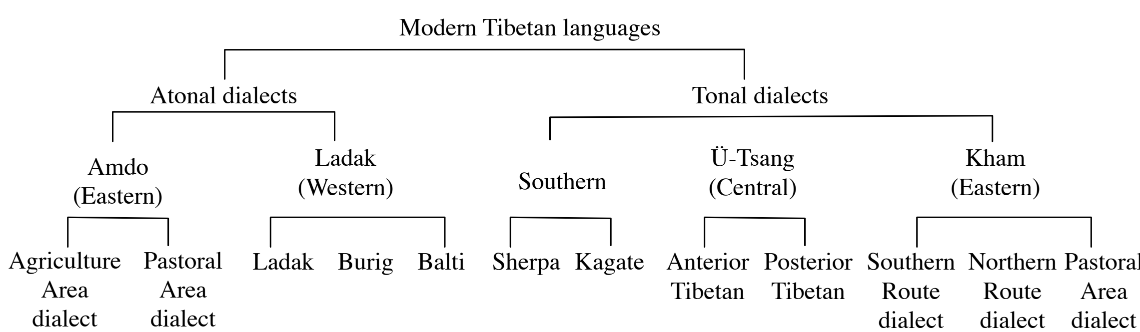


Figure 1.3: Classification of Tibetan languages based on tonal and atonal subfamilies (Hu, 2002, p.78).

1.1.1 Sociostylistic registers

LT has three socio-stylistic registers: *a*) literary *b*) religious, and *c*) vernacular.The literary register *yid-skad*, or *mkhas-pa'i kha-skad* 'speech of the learned', is used by Tibetan intellectuals and the educated class of the whole region (Tournadre and Dorje, 2003). It is more prestigious than its modern vernacular counterpart, i.e., LT or Standard Tibetan (DeLancey, 2003a; Tournadre and Dorje, 2003). Its pronunciation closely reflects the Classical Tibetan (CT) orthography, but it is seldom used in daily speech and is restricted to intellectual contexts. The religious register, or *chos-skad* 'language of Dharma,' is a variant of the literary form with identical grammar but a different lexicon, used mainly by the lamas, specifically in religious and philosophical contexts. The vernacular register is known as *bod-skad* 'Tibetan language,' *kha-skad* 'spoken language,' or *yul-skad* 'ordinary language' (Tournadre and Dorje, 2003). This vernacular variety is considered to be the standard and also acts as the *lingua franca* of the Tibetan people.

1.1.2 History of the development of Written Tibetan (WT)

Research of the historical development of modern Tibetan languages is mainly based on philological studies, on a long-established literary tradition, and on comparative historical linguistics. The latter establishes the historical development of a language by reconstructing proto-languages based on the comparison of the languages of a language family (Campbell, 2004, p.122-167). In this section I only provide a concise description of the history of the linguistic development of Tibetan literary tradition. For a general overview of Tibetan diachronic linguistics, I would like to refer my reader to the works of Sprigg (1972), Hu (2002, p.20-73), and Matisoff (2003), which I will not attempt to replicate here.

The predecessors of LT were closely associated with the development of the writing system of Tibetan, since they were the dialects spoken in the traditional seat of government, the centre of Tibetan Buddhism, and the largest metropolitan centre (Lhasa) in Tibet. Although Tibetan written records preserve important linguistic changes, it is prudent to remind the reader that the development of literary changes and colloquial changes of Tibetan languages are not always parallel (Nishida, 1970).

Based on the literary tradition, the historical development of Tibetan is commonly divided into two periods: Old Tibetan[2] (OT), and Classical Tibetan[3] (CT).

Old Tibetan was the main language spoken in the Yarlung valley which is often referred to as "the cradle of Tibetan civilization." It is the place of origin of the Tibetan Empire (600-850 AD), which was established by King Songtsen Gampo (who died 649 AD) (Denwood, 1999, p.14). This Tibetan Empire dominated the Tibetan Plateau and Central Asia from the 7th to 9th century (ibid). The Tibetan writing system, based on Old Tibetan, was introduced during the initial expansion of the Empire (650 AD)[4] to replace the original complex Sinographic-based writing system. The desire to create a new writing system was motivated by the internal communication needs of this vast kingdom and the need to translate Buddhist texts to facilitate the Tibetan people's conversion to Buddhism (Denwood, 1999; Hu, 2002; Hill, 2010). Old Tibetan continued to be the *lingua franca* for decades within the Tibetan Empire and commercial centres (e.g., Dunhuang) along the Silk Road even after the decline of the Empire and continued to be used in Buddhist texts until the post-Tibetan Empire period (mid-9th to late 10th century) (Hill,

[2]OT is also known as Bodish Tibetan (Shafer, 1968).

[3]CT is also referred to as Literary Tibetan in the literature (Goldstein et al., 1991).

[4]This is based on the year of the first entry in the Old Tibetan Annal (650-764 AD), which is the sole reliable reference on the first half of the Tibetan Empire. These manuscripts were discovered in the sealed cave library of Dunhuang in 1900 AD and were acquired by the British Museum and British Library in London and Bibliothèque Nationale de France in Paris (Dotson, 2010; van Schaik, 2010).

2010; Takeuchi, 2012).

After the introduction of the Old Tibetan writing system, the Tibetan languages continued to diversify, increasing their mutual unintelligibility. Eventually, Old Tibetan as a spoken language became extinct after the collapse of the Empire (Hill, 2010). The Old Tibetan writing system was gradually replaced by a writing standard based on Classical Tibetan (CT), which was known as the Written Tibetan (WT). According to Huang (1986), consonant cluster simplification and tonogenesis occurred during this period (12th to 13th century). Written Tibetan remains in use in the areas influenced by Tibetan culture to the present day. It can be found in both Buddhist and non-Buddhist texts. Written Tibetan, thus, is relatively stable and continues to interact with the development of modern Tibetan languages, including LT. In LT, the non-vernacular registers, both literary and religious, are heavily influenced by Written Tibetan.

1.1.3 Modern research in Tibetan

In the modern research on Tibetan languages conducted by Western scholars, one can recognize two distinctive periods: pre-exilic (1635-1959) and post-exilic (1959-Present). The initial Western investigations were carried out by Jesuit missionaries on Tibetan soil in the 17th and 18th centuries (Desideri, 2010). Many of their early works were apologetic literature, and translations of Christian literature into Tibetan languages (and vice-versa). The first Tibetan-Italian dictionary, with 35,000 entries, was compiled by Father Orazio della Penna. This dictionary was based on a manuscript of Father Ippolito Desideri, a Jesuit missionary pioneer serving in Tibet in the early 18th century (Pomplum, 2010). The most prominent scholar for the pre-exilic period, however, was Alexander Csoma de Koros, a Hungarian Buddhist monk[5] (Hetenyi, 1972), who started to learn and investigate the

[5] In Tibetan, he is also known as *Phyi-glin-gi-grwa-pa* 'the Foreign Pupil'

Tibetan language in the northwestern India-Tibet border regions. He was the first Western scholar to produce a full grammatical description and an English-Tibetan Lexicon (Koros, 1834a,b) , and he is considered to be the founder of modern Tibetology. His pioneering works and earlier missionary works in Tibet provided a viable point of access for the Western scholars that followed to carry out Tibetan studies (Jäschke, 1883; Bell, 1919; Hannah, 1912; Rörich, 1931; Koerber, 1935; Miller, 1955). In general, the grammatical descriptions during the pre-exilic period were based on the archaic Classical Tibetan written language. Linguistic information was provided by learned native informants, and analyses were presented in a typical Indo-European conjugation paradigm.

The next major period of research in Tibetan languages began after the mass exile of the Tibetan people to various diasporic communities in the Himalaya region and eventually to various European and North American countries, from 1959 onward. As a consequence, Western scholars and linguists of Tibetan have had better access to the speech community, and conversely, native Tibetan scholars have had the opportunity to acquire modern linguistic training and research techniques.

A small number of significant works established the foundation for contemporary Tibetan linguistic research. Wylie (1959) developed a romanized transcription standard that became widely adopted among Tibetologists. Chang's (1964, 1978) phonetic research and IPA-based Tibetan corpus initiated an instrumental investigation into Tibetan languages. One could say that three main areas of research caught the interest of linguists: word tonology, tonogenesis, and general morphosyntax. The former is one of the main topics currently being investigated in this dissertation. While a substantial amount of research is being conducted on the word tonology of LT, there is hardly any research on the intonational phonology of LT. Previous research in LT word tonology and intonational phonology will be further reviewed in §2.1 and §2.3, respectively.

1.2 Phonological and morphological sketch of LT

In this section, I will provide a basic description of the segmental inventory, syllable structure, and morphology of LT. The segmental inventory and syllable structure provide important insights into LT tonogenesis (§1.3). The morphological description provides background information required to understand the interaction of morphology and phonology: Morphology is key to interpreting the pitch patterns of morphologically complex words and phrases in LT. The incorporation of the prosodic structure into the study of tone and intonation is one of the major research objectives of this dissertation. In this morphology section, I provide a brief description of the morphology of nouns, verbs, and adjectives.

1.2.1 Segmental inventory

There is significant disagreement regarding LT segmental inventory. The earliest descriptions, as well as some recent studies, are mainly orthography-based (Koros, 1834a; Jäschke, 1881; Hannah, 1912; Bell, 1919; Sedláček, 1959; Goldstein et al., 2001). These analyses proposed a segmental inventory that largely resembles Classical Tibetan. More recent descriptions adopt a mostly phonetic or a phonemic approach (Hari, 1977; Dawson, 1980; Hu, 2002; DeLancey, 2003b; Denwood, 1999; Tournadre and Dorje, 2003). The phonemic inventory of consonants and vowels of LT is given in Table 1.2 and 1.3, respectively (cf. Hu, 2002, p.219). It is important to note that onset obstruents have lost voicing contrast in colloquial speech. However, voiceless plosives (p, t, k) are often voiced when a syllable bears low tone in citation form and intervocalically (cf. Lim, 2010b). This could be due to the fact that some of my participants are learned Tibetans who still preserve the voicing contrast according to the Classical or Written Tibetan, but is more likely due to a subphonemic redundancy between voicing and the low tones, and crosslinguistic tendency to intervocalic voicing.

Consonants	Labial	Alveolar	Retroflex	Palatal	Palatal-ized velar	Velar	Glottal
Plosive	p p^h (b)	t t^h (d)	ʈ $ʈ^h$ (ɖ)		k^j k^{hj} (g^j)	k k^h (g)	
Affricate		ts ts^h		tɕ $tɕ^h$			
Nasals	m	n		ɲ		ŋ	
Trill		r					
Fricative		s (z)		ɕ/ʃ (ʑ)			h
Approximate	w			j			
Lateral		l ɬ					

Table 1.2: Consonant inventory of LT.

Vowels	Front		Back
	Unrounded	Rounded	Rounded
Close	i	y	u
Mid	e	ø	o
Open	ɛ		ɑ

Table 1.3: Vowel inventory of LT.

1.2.2 Transcription of Written Tibetan

Written Tibetan (WT) has thirty basic graphemes. It is an abugida, or an alphabetic system in which all vowels are but one indicated by using diacritics (Rogers, 2005). In WT, the vowels [i, e, o, u] are indicated with diacritics and the vowel [a] is unmarked. It is important to point out that the Tibetan orthography is based on 11th century Classical Tibetan, while the pronunciation of LT along with that of other Tibetan languages have continued to evolve independently of the orthography (Denwood, 1999, p.40-41). Thus, there is no tone marking in WT since tonal contrast was not developed when it as designed. In general, I will give the Wylie transliterations for WT (Wylie, 1959) along with IPA transcriptions of modern Lhasa Tibetan, as it is the standard among Tibetan researchers, but the lay reader can safely ignore them. Wylie transliterations will be presented in *italic*. The transliterations of the basic graphemes in the traditional order are illustrated in Figure 1.4.

ཀ་	*ka*	ཁ་	*kha*	ག་	*ga*	ང་	*nga*
ཅ་	*ca*	ཆ་	*cha*	ཇ་	*ja*	ཉ་	*nya*
ཏ་	*ta*	ཐ་	*tha*	ད་	*da*	ན་	*na*
པ་	*pa*	ཕ་	*pha*	བ་	*ba*	མ་	*ma*
ཙ་	*tsa*	ཚ་	*tsha*	ཛ་	*dza*	ཝ་	*wa*
ཞ་	*zha*	ཟ་	*za*	འ་	*'a*	ཡ་	*ya*
ར་	*ra*	ལ་	*la*	ཤ་	*sha*	ས་	*sa*
ཧ་	*ha*	ཨ་	*a*				
ཨི	*i*	ཨུ	*u*	ཨེ	*e*	ཨོ	*o*

Figure 1.4: The Tibetan abugida and its Wylie transliteration (Tournadre and Dorje, 2003).

1.2.3 Syllable structure

LT has the simple syllable structure as shown below:

(1) $(C_i)V(:)(C_f)$

All consonants can be found in the syllable-initial position (C_i), but there is a restricted set of syllable-final consonants (C_f). As previous authors, I consider that there are no onset consonant clusters because palatalized velars [k^j k^{hj}] behave as unitary segments (DeLancey, 2003b; Tournadre and Dorje, 2003; Goldstein et al., 1991). The complete list of segments possible in each position is shown in Table 1.4. Lastly, the syllable is the tone-bearing unit of LT and its role in word tonology will be further discussed in Chapter 3.

Position	Possible Segment
C_i	All consonants
V	All vowels
C_f	p, k, ʔ, m, ŋ, n, r

Table 1.4: Phonotactics of LT.

1.2.4 Morphology

LT is primarily a monosyllabic language like many of its Sino-Tibetan relatives. In this section, I will provide a concise morphological description of its nouns and verbs.

Nouns in LT consist primarily of monosyllabic lone stems. Monosyllabic nouns are the most archaic items of the Tibetan lexicon, and many of them have cognates in other Sino-Tibetan languages (Hu, 2002). There is also a limited quantity of true polysyllabic noun stems, usually not more than trisyllabic (Qu and Tan, 1983; Denwood, 1999). Furthermore, LT uses affixation and compounding as strategies to generate morphologically complex words. The former morphological operation generates derived nouns, the latter, compound nouns. The majority of these words are disyllabic, and usually not more than quadrisyllabic (cf. Kong, 2012). In what follows, I will provide a concise overview of the morphology of three types of noun stems: simple nouns, derived nouns, and compound nouns.

Simple nouns are usually monomorphemic, and there are of two types: simple forms and subtractive forms. Examples of both noun forms are included in Table 1.5. The subtractive forms might appear synchronically monomorphemic, but

Simple nouns (monomorphemic)					
lha	[ɬá]	'god'	*mu tig*	[mùtí]	'pearl'
mi	[mǐ]	'person'	*mu ge*	[mùkí]	'famine'
sa	[sá]	'earth'	*gzhis ka rtse*	[ɕ̥ìkátsé]	'Shigatse'
bod	[pʰø̌ʔ]	'Tibet'	*katora*	[kátórà]	'red copper box'
deb	[tèb]	'book'	*'bi lar sa*	[pìlásá]	'cement'
rgun 'brum	[kṹːtʂṹː]	'grapes'			
Subtractive nouns					
koba	[kóː]	'cowhide'	CT: *koba* →LT: *kóː*		
glaba	[láː]	'musk deer'	CT: *glaba*→LT: *láː*		
lbuba	[pǒː]	'bubble'	CT: *lbuba*→ LT: *pǒː*		

Table 1.5: Examples of simple and substractive nouns (Qu and Tan, 1983).

they were derived from noun stems with nominalizer suffixes whose onsets were deleted. The original constructions were preserved in Old Tibetan or Classical Tibetan written forms. A number of monosyllabic noun stems originated from this process in LT, but they are not as prolific as in other Tibetan dialects (Qu and Tan, 1983).

Derived nouns have two types of affixation: prefixation and suffixation. There exists only one prefix, [a-], which attaches to a small number of noun stems denoting kinship terms (Qu and Tan, 1983), as shown in Table 1.5. The most common type of derived nouns is verb stems followed by a nominalized suffix. There are, however, many suffixes that can be attached to the noun stem. The functional load of these nominal suffixes has significantly diminished in the past centuries; some scholars consider them to be 'synchronically meaningless' (DeLancey, 2003b). Examples of derived nouns are given in Table A.1 of Appendix A.

There are numerous compound nouns in LT (Goldstein et al., 1991; Qu and Tan, 1983; Denwood, 1999; DeLancey, 2003b). The majority of the compound nouns are either disyllabic or quadrisyllabic, as trisyllabic compound nouns are relatively scarce. A compound noun can consist of more than one noun stem, but

Prefix		Stem		Derived noun		
a	/á-/	*zhang*	/ɕáː/ ‘uncle(maternal)’	*azhang*	[áɕáː]	‘uncle (maternal)’
a	/á-/	*ne*	/ní/ ‘aunt (paternal)’	*ane*	[ání]	‘aunt (parternal)’
a	/á-/	*cag*	/tɕáʔ/ ‘older sister’	*acag*	[átɕáʔ]	‘older sister’

Figure 1.5: Derived nouns of LT: prefix-stem.

it can also be comprised of stems from other grammatical categories. Details of possible combinations are presented in Table A.2 of Appendix A.

LT verbs utilize compounding and affixation to generate morphologically complex verb words and phrases. Affixation in LT verbs is especially prolific and complex. Verbs are the only obligatory constituent in a sentence. Figure 1.6 shows the basic verb morphological template.

		TAM_A			(TAM_B)	
(NEG_2)	V_{stem}	(CP_1)	(NEG_1)	TAM_1	(NEG_2)	(TAM_2)

Figure 1.6: Basic morphological templates of LT verb phrase.

In the following paragraphs, I give a brief overview of the morphology of the verb stem, which may consist of a simple verb, compound verb, or serial verb. I will then provide a concise discussion of affixation, which includes the negation marker and Tense-Aspect-Modality (TAM) suffixes. I will focus only on TAM_A suffixes which are more common. Probability TAM suffixes (TAM_B) are excluded from the discussion due to insufficient data.

There is a relatively limited number of simple verbs in LT. These simple verbs are mainly monosyllabic and represent a closed class. Duff (2005, p.584) estimates that there are approximately 1200 of them. LT verbs are categorized according to transitivity and volition (Agha, 1993; Tournadre and Dorje, 2003). Examples of these verbs are shown in Table 1.6. Historically, most verbs had four distinct stem forms: past, present, future, and imperative. Now the stem forms are highly irregular and unpredictable in LT (Denwood, 1999, p.105-107).

	Volitional			Non-volitional		
Transitive	*bltas*	[tá]	‘to look’	*mthong*	[tʰóŋ]	‘to see’
	nyan	[ɲøn]	‘to listen’	*go*	[kʰò]	‘to hear’
	dbral	[bʲàl]	‘to tear’	*ral*	[ràl]	‘to get torn’
Intransitive	*snyol*	[ɲø̀ː]	‘to lie down’	*na*	[nà]	‘to be sick’
	langs	[làŋ]	‘to get up’	*shi*	[ɕì]	‘to die’
	’gro	[ʈò]	‘go’	*drag*	[ʈʰà]	‘to recover’

Table 1.6: Examples of simple verb categories (adapted from Tournadre and Dorje (2003, p.141).)

Compound verb stems, also known as phrasal verbs (Agha, 1993, p.105), are prevalent in LT. Compound verbs normally consist of a noun root followed by a verbalizer. They are different from serial verbs, in which two verb stems are in sequence ($[V_1V_2]_{stem}$). The three most common verbalizers are: *a*) *kya* [kʲà] ‘to make’ , *b*) *btang* [táŋ] ‘to send, to do’, and *c*) *byed* [tɕʰèʔ] ‘to act’. Examples of verbalized nouns can be found in Table A.3 in Appendix A or refer to the work of Bartee (2000) which has a more extended list of verbalizers in LT.

The second verb, called a secondary verb (V_2) in a serial verb, is also known as a ‘light’ verb or compound verb in some descriptions of Tibetan grammar (DeLancey, 1991). It is picked from a small closed class; see Table A.4 of Appendix A for more details.

TAM suffixes are elements that provide essential temporal, evidential modal, and aspectual information (DeLancey, 1992; Denwood, 1999; Tournadre and Dorje, 2003). In LT, there are four categories of tense/aspect: *a*) future, *b*) present, *c*) imperfective past, and *d*) perfective past. In each category, there are three types of evidentiality[6]: *a*) ego (EGO), *b*) indirect (IDIR), and *c*) direct (DIR). The TAM

[6]The first evidential category, ego, indicates the speaker has privileged access to the information to construct his/her proposition. This is an evidential category distinctive of the Tibeto-Burman language family (Garrett, 2001). Garrett (2001, p.11) observes that the category of ego (EGO) evidentiality is associated with the TAM suffixes [jin] and [jøʔ]. In addition, the indirect evidentiality (IDIR) category is associated with the TAM suffixes of [reʔ] and [joːreʔ]. There is a dichotomy be-

suffixes are the most intricate constituents of the predicate structure of LT. TAM suffixes are often a portmanteau, as they encode two or more TAM categories at any single morphological slot. Given the morphological facts, it seems clear that we need to recognize a template with two distinct positions, one occupied by the connecting particles (CP) *ki* [-ki] and *pa* [-pa], which may sometimes be absent (∅), and another associated with the TAM suffixes proper: *yin* [-jin], *red* [-reʔ], *yod* [-jø], *yod red* [joːreʔ], *'dug* [duʔ], and *song* [soŋ] (fused form). I summarize the proposed morphological structure for LT TAM suffixes in Table 1.7, where the connecting particles and the TAM suffixes are placed in two separate columns (cf. Tournadre and Dorje, 2003, p.462)[7]. The negative form of these TAM suffixes can be found in Table 1.5 of Appendix A.

1.3 Tonogenesis in LT

Tonogenesis has been a key area of research interest since linguists began investigating tonal phenomena in Tibetan languages. Understanding this process is crucial to grasping the tonal and intonational phonology of innovative tonal Tibetan languages such as LT. In what follows, I will briefly discuss tonogenesis in general, the views of traditional grammarians on 'tone', and will then proceed to review various research results concerning tonogenesis.

Modern tonogenetic research began with an investigation into the origin of Vietnamese tones. Based on a historical comparison of Sino-Vietnamese loan words,

tween first person and non-first person TAM suffixes. The ego TAM suffixes are normally required if the subject of the clause is the first person. The direct and indirect TAM suffixes are normally required when the subject is the second or third person. As Denwood (1999, p.135) demonstrates, direct and indirect TAM suffixes can be used with the first person subject to indicate involuntariness. For a more detailed description of the epistemics and evidentials of LT, please refer to the to be appear article of Caplow (2016).

[7] Some verbs might preserve the four distinct stem forms with respect to the tense/aspect, others merged irregularly. That said, the present verb stems tend to associate with future and present tense; the past verb stems tend to associate with imperfective and perfective past.

			TAM_A	
Tense/aspect	V	Evidential	CP_1	TAM_1
Future	$V_{(pres)}$	EGO	*ki* [-ki]	*yin* [-jin]
		IDIR		*red* [-reʔ]
		DIR	N/A	
Present	$V_{(pres)}$	EGO	*ki* [-ki]	*yod* [-jøʔ]
		IDIR		*yod red* [-joːreʔ]
		DIR		*'dug* [-duʔ]
Imperfective Past	$V_{(past)}$	EGO	*pa* [-pa]	*yin* [-jin]
		IDIR		*red* [-reʔ]
		DIR	*song* [-soŋ]	
Perfective Past	$V_{(past)}$	EGO	$\varnothing$	*yod* [-jøʔ]
		IDIR		*yod red* [-joːreʔ]
		DIR		*'dug* [-duʔ]

Table 1.7: TAM suffix and connecting particles.

Maspero (1912) identified that Vietnamese tones split into two groups based on the voicing of the word-initial consonants (cf. Ferlus, 2004, p.297). This was further investigated in André G. Haudricourt's (1954) seminal work, in which he found that three tones were derived from archaic Vietnamese laryngeal codas (Haudricourt, 1954). This model of tonogenesis is also applicable to Thai and Karen (Haudricourt, 1961). Haudricourt's model was later adopted by Matisoff (1973) and applied to a vast range of tonogenetic phenomena in East and Southeast Asian languages.

It is the current consensus that the intrinsic phonetic characteristics of laryngeal gestures that induce F0 perturbations are the source of the emergence or multiplication of tones (Matisoff, 1973; Purcell, 1974; Hyman, 1976; Mazaudon, 1977; Purcell et al., 1978; Hombert et al., 1979; Svantesson, 1989; Thurgood, 2002; Sun, 2003; Kingston, 2005; Hyslop, 2009; Kingston, 2011; Hyman, 2013). The phonologization of these intrinsic phonetic characteristics is a "compensatory mechanism for the depletion of consonantal contrast at syllable-final or syllable-initial position" (Matisoff, 2000, p.86). These claims are supported by insights from instrumental studies of tonal languages of Southeast Asia (Thurgood, 2002; Brunelle et al., 2010). According to Kingston (2011, p.2304-2327), the following are known sources of tonogenesis:

1. *From preceding consonants*: Concomitant with a decrease or total loss of phonation contrast of onsets, the intrinsic F0 associated with the onset's phonation is phonologized. For example, the contrast between an intrinsic high F0 of a voiceless onset obstruent and an intrinsic low F0 of a voiced onset obstruent is transferred to the F0 contrast between a high and low tone after the loss of voicing contrasts.

2. *From following consonants*: An automatic F0 perturbation induced by the coda is phonologized. For example, a final obstruent that generates a rising pitch contour will encourage the development of a rising tone, and a final voiceless fricative that generates a falling pitch contour will encourage the development of a falling tone. However, an open and nasal coda that preserves the original level pitch contour will encourage the development of a level tone.

3. *From tone splitting*: Tone splits due to onsets are a common source of tonogenesis in tonal languages found in East and Southeast Asia. This is mainly attributed to the phonation characteristics of onset consonants.

4. *From uncommon sources*: Tonal contrasts can originate from the intrinsic F0, which is associated with vowel height. This is found in Hu. A high vowel, which has higher intrinsic F0, will develop into a high tone, and non-high vowel, which has lower intrinsic F0, will acquire a low tone. The advanced tongue root feature is another unusual source for new tonal contrasts. [+ATR] vowels have higher intrinsic F0 than [-ATR] vowels. This feature is responsible for the contrast between the high and extra high tones in Western Lugbara.

5. *From stress and intonation to tone*: The tone originates from the F0 correlate of stress, as found in North Germanic languages such as Swedish, Norwegian, and Danish.

1.3.1 Tone in traditional Tibetan grammar

Traditional Tibetan grammarians consider Tibetan to be atonal (Duff, 2005, p.404). This is not surprising because the writing system predated tonogenesis, and there is no explicit tone marking on the Tibetan script. That being said, they recognized 'tone' as an integral part of correct pronunciation. *The Root of Grammar: The Thirty Verses*, a seventh-century traditional grammar description, defines the role of tones in literary training:

> Persons wanting to put effort into a discipline
> First must learn the tones then
> The threefold prefixes, name-bases, and suffixes
> Will be their training for reading.
> (Duff, 2005, p.145)

In traditional Tibetan grammar, these 'tones' are referred to as *dbyangs*. They are not exclusively used to describe the pitch of a given rhyme. 'Tone' is also associated with the acoustic perturbation introduced by the voicing of the consonant on the onset of the vowel. It is traditionally described as the 'roar of a consonant,' and the 'vowel is seen as the tone that flies as a melody over the base sound of a consonant' (Duff, 2005, p.260). From that point of view, the intrinsic pitch or 'tone' associated with the phonation of the onset consonant is peripheral and without lexical status, but it is required for correct pronunciation. For example, the perceived pitch of the rhyme that follows a voiceless onset stop (*k*) is described as follows: 'it feels as though it is high in the head and palate' (Duff, 2005, p.402). If it follows an aspirated voiceless onset stop, e.g., $k^h a$, it is described as having 'a softer quality than *ka*' (Duff, 2005, p.403). However, if it follows a voiced onset stop, e.g. *ga*, it is described as 'being made with a lower tone than the first letter *k*' (Duff, 2005, p.403).

The descriptions above support the idea that the phonemic tones of modern Tibetan languages were derived from earlier phonation contrasts (Matisoff, 2000; Hu, 2002). In other words, the intrinsic pitch associated with the phonation of a given stop was phonologized and used as phonemic tone. I will further discuss this phonologization process in the following sections.

1.3.2 General path to tonogenesis in LT

Research on Tibetan languages concludes that tonogenesis in Tibetan languages is associated with the simplification of onset clusters and loss of voicing contrast of onset obstruents (Matisoff, 2000; Hu, 2002). Although the role of the former is open to discussion, the latter is clearly the main driving force. This conclusion is deduced from the comparison of written texts from various eras, and historical comparative studies on closely related languages. According to the Tibetan tonality continuum proposed by Sun (1997), LT is at an advanced tonogenetic stage, as shown in Table 1.8 as contrasts in pitch levels (i.e., high versus low) have become phonemic. The shape of pitch contours (i.e., falling versus rising), however, is not yet contrastive. In the following paragraphs, I will present the currently established consensus regarding the tonogenesis path of LT.

Tonality scale	Description of each stage	Representative dialects
Atonal	no phonemic tone or redundant 'habitual' tone	Ndzorge, Ngaba
↓	no phonemic tone; redundant 'habitual' tone	Labrang, Zhangia
↓	tone phonemic in restricted environment only	Amdo, Sherpa, Balti
↓	tone generally phonemic, tone values unstable/non contrastive in some syllable types	Derge, Chamdo
↓	stable tone values; contrasting pitch level but high redundancy	Lhasa, Gar
Tonal	additional contrast between level and falling pitch contours	Shigatse, Dzongkha

Table 1.8: The Tibetan tonality continuum (Sun, 1997, p.487).

	Stage I	Stage II	Stage III
Simplification	$\#C_1C_2C_3$- $\#C_3$-	$\#C_2C_3$- $\#C_3$-	$\#C_3$-
Phonemic contrast (voicing)	+	+	-
Tonal contrast	-	(+)	+

Table 1.9: Tonogenesis of LT (Nishida, 1975, p.48).

According to Nishida (1975), LT tonogenesis can be divided into three different stages, as shown in Table 1.9. At the original stage (or Stage I) before tonogenesis, there were two distinctive types of onset clusters: complex ($\#(C_1)C_2C_3$) and simple ($\#C_3$). The voicing contrast of the onset obstruents ($\#C_3$) were intact and there was no tonal contrast. At Stage II, the complex onset clusters were simplified ($\#C_2C_3$) in colloquial speech and the intrinsic F0 associated with the voicing of the onset obstruents became extrinsic (cf. Hyman and Magaji, 1970). That said, these complex onset clusters were still preserved in the writing system. At Stage III, there were no more complex onset clusters, and the voicing contrast of onset obstruents was lost. Subsequently, the extrinsic F0 associated with the voicing contrast of onset obstruents became phonemic, i.e., tonal contrast. In Central Tibetan languages such as LT, Stage II took place between the 8th and 9th centuries, and Stage III occurred in the 10th-century (Nishida, 1975, p.49).

The origin of tonal contrast in LT can be further illustrated by the distribution of tones based on the onsets and codas of Written Tibetan (cf. Nishida, 1975; Sedláček, 1959). As mentioned earlier, Written Tibetan predates tonogenesis. I have adapted Nishida's findings in the following Table 1.10[8].

		Onset	
		$C_3 =$ nasals/w/j/l/r	
		$C_3 =$ vl. obstruents	$C_3 =$ vd. obstruents
Coda	-m/-n/-ŋ/-r/-l/-s	high-level	low-rising
	-b/-g	high-falling	low-level

Table 1.10: Distribution of allotones of monosyllabic words of LT based on partial historical onsets and codas of written Tibetan.

According to Nishida, the voicing of historical onset obstruents (C_3) determines the pitch level of the rhymes: voiceless onset obstruents correspond to the high-onset allotones (i.e., high-falling and high-level), while the voiced onset obstruents correspond to the low-onset allotones (i.e., low-level and low-rising). The words with sonorant onsets (nasals/w/j/l/r) could go in either direction. Furthermore, the historical codas determine the shape of the pitch contour. Consequently, codas of words dictate the rhyme duration, which affects the pitch realization of tones. Sonorant codas (-m/-n/-ŋ/-r/-l/-s) yield longer duration, while words with obstruents codas [-b/-g] or the absence of codas result in shorter duration, which indirectly yield different allotones.

[8]Nishida proposed a four-tone analysis: high-level, high-falling, low-rising, and high-rising. The latter two are slightly different from the conventional four-tone analysis which are usually analyzed as low-level and low-rising, respectively (see §2.1.2). In this dissertation, all the high-onset allotones are analyzed as underlyingly an /H/, and all the low-onset allotones are analyzed as /LH/ (see Chapter 3).

1.3.3 Recent studies on Tibetan tonogenesis

Two recent studies have suggested additions and revisions to Nishida's model of Tibetan tonogenesis. Caplow (2009) discusses the role of stress patterns in tonogenesis in Tibetan languages while Kong (2012) proposes utilizing homonym rates to determine the stages of the tonogenesis in Tibetan languages.

1.3.3.1 From stress to tone

A number of researchers have proposed that stress played an important role in Tibetan tonogenesis (Goldsmith, 1980; Bielmeier, 1988; Caplow, 2009). In this section, I will focus on Caplow's (2009) reconstruction of the Proto-Tibetan stress patterns in disyllabic words based on two archaic and atonal Tibetan languages (Balti and Rebkong Amdo). The reconstructed stress pattern for non-verbal words is *σσ́, where the second syllable is stressed. The verbs have a contrasting stress pattern *σ́σ, where the first syllable is stressed. The primary phonetic cue for the stressed syllable is pitch, where the stressed syllable has a higher pitch than the others. Caplow focuses on the reconstructed stress pattern of non-verbal words and argues that this acoustic template has been preserved through diachronic changes. She claims that tones in tonal Tibetan languages originated from this stress pattern because it corresponds to the most common tonal pattern of these innovative Tibetan languages: the tone is always contrastive in the first syllable (low versus high register), but it is consistently a high tone on the second syllable of disyllabic words (Caplow, 2009, p.527-528). Similar (though not identical) tonogenetic paths have been proposed for Scandinavian languages such as Swedish and Norwegian (Riad, 2013).

Caplow's proposal is important in that it seems to explain the synchronic tonal contrast in modern LT disyllables. At an earlier stage, LT disyllables would have had a stressed-conditioned non-contrastive LH contour. The loss of voicing and the tonogenesis deriving from it would have resulted in tone split in the initial syllable, yielding LH vs. HH.

That being said, Caplow's model also has limitations. First, it cannot currently explain the tone patterns of non-disyllabic words (but could possibly be extended to capture them). More seriously, it incorrectly predicts that verbs should have different tonal patterns than others words in LT, as verbs in atonal Tibetan languages typically have a word-initial stress, rather than word-final stress. In addition, most researchers agree that verbs have the same tone patterns as non-verbs in LT (Chang and Chang, 1964; Sprigg, 1955; Goldstein and Narkyid, 1984; Denwood, 1999; Hu, 2002; Tournadre and Dorje, 2003).

1.3.3.2 Homonym rate and tonogenesis

Kong (2012) tested the claim made by Nishida (1975) that reduction of consonant clusters in Tibetan languages and the associated increase in homonymy may have been a causal factor behind tonogenesis. In his paper, the homonym rate (HR) was calculated based on the corpora of different Tibetan "dialects" (Xiahe, Zeku, Batang, Dege, Rikaze, and Lhasa) and compared with Written Tibetan (WT) or Classical Tibetan (CT), which has a homonym rate of zero.

Kong calculates two types of homonym rates (HR): HR1[9] and HR2. The former represents the homonym rate based on the onsets and codas of monosyllabic words, as shown in the following formula:

$$HR1 = \frac{TN}{NIF} - 1 \quad (1.1)$$

[9]Kong also proposed that based on HR1, one would be able to interpolate the year that the tonal contrast was established in Tibetan Languages. I, however, remain skeptical about this particular extrapolation, because he assumes the homonym rate is constant.

TN represents the total number of words in the written Tibetan corpus, and NIF[10] represents the number of words in the current Tibetan language with different onsets and codas.

HR2, on the other hand, represents the homonym rate of monosyllabic words when all segments are taken into account, as shown in the following formula:

$$HR2 = \frac{TN}{NS} - 1 \tag{1.2}$$

The definition of TN is the same as above, but NS represents the number of unique monosyllabic words based on all their segments. Thus, higher HR1 or HR2 value means higher homonym rate. Kong provides the summary of what happens in six Tibetan dialects in the followings:

Dialect	HR1	HR2
WT	0	0
Xiahe	1.9089	1.9088
Zeku	2.5776	2.5775
Batang	3.0408	2.2096
Dege	3.2283	1.9873
Rikaze	4.2017	1.6521
Lhasa	4.3698	1.7263

Table 1.11: Homonymic rate based on initials and finals.

Overall, there seems to be a strong correlation between HR1 and tonogenesis: Lhasa, or LT, has the most advanced tonogenesis among the six selected dialects, while Xiahe and Zeku have the least advanced tonogenesis. Interestingly, the match is not as good when vowels are included in the calculation of homonymy (i.e., HR2). Thus, even if the mechanisms are unclear, there seems to be some grounds to the claim that cluster simplification played a role in Tibetan tonogen-

[10] Number of initials and finals

esis.

1.4 Organization

This dissertation is organized as follows. In Chapter 2, I present a review of the literature on synchronic tone, stress, and intonation in LT. In Chapter 3, I focus on word-level tonal phonology and present experimental results on the surface realization of LT words and the status of word stress. In Chapter 4, I present the intonational phonology of LT. This includes the integration of prosodic structure and the interaction between lexical tones and intonational tones. In Chapter 5, I develop an F0 synthesis model to validate the adequacy of the proposed tonal and intonational phonology of LT. In Chapter 6, I present my conclusion and plans for future research.

Chapter 2

Previous linguistic research on word tone, stress, and intonation in LT

In this chapter, I provide an overview of previous linguistic research on word tone (§2.1), word stress (§2.2), and intonation (§2.3). These reviews provide necessary background information to facilitate the discussion in the latter chapters.

2.1 Previous research on word tone

Previous work has not reached a consensus concerning the inventory of tones in LT, which can range from zero (Civera, 1970) to eight (Fang Hu, 2010). It is staggering that this disparity is based on observations of the pitch pattern of the same language.

A few factors contribute to this controversy concerning the tonal inventory of LT. One of them is the failure to differentiate between lexical tones and allotones. A lexical tone can potentially have different types of allotones depending on the context. This inadequacy of classification often unnecessarily inflates the inventory of tones. Furthermore, most earlier researchers of LT did not account for the interaction between lexical tones and prosodic structure (Sprigg, 1955; Hu, 2002; Monich, 2011). Similarly, the interaction between lexical and intonational tones are not taken into consideration. In this section, I will present the analyses proposed by Sprigg (1993), Monich (2011), and Gong (2014). I will then compared them to my own analysis in Chapter 3.

In the Tibetan linguistic literature, the term *tone* can represent one of the following concepts: *a*) *lexical tone*, *b*) *allotones*, or *c*) *pitch contour*. To avoid confusion, I will provide brief definitions of the above before proceeding to further discussion.

Lexical tones are also known as phonemic tones or tonemes in the literature. A lexical tone is an underlying and abstract pitch pattern associated with a tone-bearing unit (syllable or mora) that can differentiate the meaning of words or grammatical functions in a given language. An LT example would be the contrast in lexical tones, between [mǎ] 'sore' and [má] 'mother'. These two words contrast because the first bears an underlying /LH/ lexical tone while the second bears an underlying /H/ lexical tone.

A lexical tone might have a different phonological output in different phonological contexts. These phonological outputs might appear to have different pitch contours, but they are in fact *allotones* of a particular lexical tone or toneme. In LT, tones can be spread over syllables and a /LH/ tone can be realized as L in disyllabic or polysyllabic words, due to the decomposition of the /LH/ to [L] and [H] tones (see Chapter 3). For example, [mè.pór] 'fire pan', consists of two morphemes, /mě/ 'fire' and /pór/ 'pan'. The first morpheme, /mě/, has an underlying /LH/ and the second morpheme, /pór/, has an underlying representation /H/. In a non-compound form, /mě/ has a phonological output [LH] realized as a low-level pitch contour, and/pór/ has a phonological output of [H] realized as a high-level pitch contour. However, as a disyllabic word, [mè.pór] has a phonological output of [L.H], realized as a low-level and high-level pitch contour, sequentially. The /H/ of the non-initial syllable is deleted, the /LH/ of the initial syllable is decomposed into /L/ and /H/, and the latter spreads to the second syllable. Therefore, the phonological outputs of [L] and [LH] are allotones of the same lexical tone /LH/. Consequently, it is impossible to determine the tonal inventory and its nature without understanding the phonological processes involved in a given language.

Pitch contour is defined as the actual F0 signal that can be perceived or extracted from the speech produced by a speaker. This pitch contour is often extracted as an F0 tracing in speech analysis software such as PRAAT (Weenink and David, 2017). For example, the pitch contour of [mǎ] *rma*'sore' is low-level, and [mǎr] *dmar* 'red' is rising. Although these two words share the same underlying representation (/LH/), they differ in pitch realization due to difference in rhyme (nucleus + coda) duration length. An /LH/ found in a word with a shorter rhyme duration is realized as a level pitch contour, while a longer rhyme duration is realized as a rising pitch contour. It is important to recognize that both lexical tones and allotones are phonologically abstract representations that may or may not be directly reflected by the pitch contour or F0 tracings. Thus, it is imperative that a researcher does not take the pitch contour at face value and "**read phonological tones off F0 tracings**" (Michaud and Vaissière, 2015). Like pitch, pitch contour can be influenced by clausal intonation, segmental effects, and the idiosyncratic nature of the individual speaker.

One should also avoid importing segmental elements (e.g., duration, coda segment) into the classification of a lexical tone. These elements are necessary for pitch pattern description, but they unnecessarily complicate abstraction of the lexical tone. Without this discernment, a researcher will unavoidably inflate the number of 'tones' in the inventory. Furthermore, minimal pairs (same segments but different lexical tones) are needed to support the contrastiveness of lexical tones.

Finally, it is imperative that researchers include the tonal pattern of polysyllabic words along with monosyllabic and disyllabic words in order to produce an accurate description and analysis of the LT tonal phenomena.

With the above criteria in mind, both the atonal analysis and the analysis claiming three contrasting pitch levels (/L, M, H/) are discarded (Civera, 1970; Bell, 1919)[1]. Similarly, if we clearly distinguish lexical tones, allotones, and surface realization, (5[2]/6/8) are just variants of the two-tone analysis. In what follows, I will focus on two representative and currently leading analyses: the basic two-tone and basic four-tone analyses. Then, I will discuss the extension of these analyses to polysyllabic words.

2.1.1 Two-tone analyses

In this section, I will present the basic two-tones analysis and the "junction" analysis approach of Sprigg (1955).

The basic two-tone analysis recognizes high and low lexical tones (Jäschke, 1881; Cao, 1930; Thonden, 1986; Yu and Chao, 1930; Qu, 1981; Ossorio, 1982; Duanmu, 1990; Sprigg, 1993; Yip, 1993, 2002; Monich, 2011; Lim, 2010a, 2012). These lexical tones have different surface realizations depending on the segments of a given word. In LT, the duration of the rhyme is a key factor that affects surface realization. The rhyme is considered to have a long duration when it contains a long vowel (V:), a vowel with a sonorant coda (V[+son]), or both (V:[+son]). Similarly, the rhyme is considered to have a short duration where there is a short vowel, and there is either no coda (V) or the coda is not sonorant (V[-son]). Thus, there are four basic types of pitch realization: *a*) high falling pitch contour on

[1] In this dissertation, I proposed that there are two lexical tones found on stems of LT (LH, H), and an additional lexical tone found on affixes of LT (L). According to current work, affixes of LT do not bear /LH/. That said, there are only two contrasting pitch levels (see Chapter 3).

[2] Five tones analysis is a four-tone analysis with the addition of neutral tone (Sedláček, 1959).

a short duration rhyme; *b*) high-level pitch contour on a long duration rhyme; *c*) low-level pitch contour on a short duration rhyme; and *d*) low rising pitch contour on a long duration rhyme (see Table 2.1).

	Short	**Long**
H	High-falling	High-level
L	Low-level	Low-rising

Table 2.1: Pitch realization of H and L lexical tones with respect to rhyme duration

Some researchers claim that there are additional patterns (up to eight). These are due to further distinguishing the F0 effect of various postvocalic consonants (e.g., l, r, h, ʔ) in conjunction with two contrasting pitch levels (Lim, 2012; Hu, 1980; Hu et al., 1982; Fang Hu, 2010).

There is virtually unanimous consensus among Tibetan tonologists supporting the underlying nature of the high lexical tone, usually labeled as 'H'. The low lexical tone, however, is labeled as either 'L' or 'LH' depending on the analysis (see Chapter 3). The examples in Table 2.2 illustrate a typical two-tone analysis with four different pitch realizations.

ka^{53}	/H/	*bka'*	'order'	$kaʔ^{53}$	/H/	*bkag*	'against'
$ka:^{55}$	/H/	*ka.ba*	'pillar'				
ka^{33}	/LH/	*sga*	'saddle'	$kaʔ^{35}$	/LH/	*'gags*	'plugged'
$ka:^{15}$	/LH/	*bsgar*	'install'				

Table 2.2: Typical two-tone analysis (Hu, 2002, p.171).

Now that we have reviewed the basic two-tone analysis of LT, we will address another variant of the two-tone analysis, namely, Sprigg's "junction" analysis approach (Sprigg, 1955, 1993). The main distinction between his approach and a plain two-tone approach is that the former refuses to analyze the tone of a word in a citation form but rather always does it in spoken sentences. As "the role of intonation in Lhasa Tibetan, therefore, is far from negligible" (Sprigg, 1993, p.476), he recognizes two lexical tones: Tone 1 and Tone 2. Tone 1 consists of the high lexical tone, while Tone 2 consists of the low lexical tone. According to this analysis, pitch level (i.e., high vs. low) is contrastive, but pitch contours (i.e., falling vs. rising) are not (Sprigg, 1993, p.500).

I reproduce two sets (a and b) of his examples in Figure 2.1. Each set consists of tonal patterns for three sequential trisyllabic verb phrases ($V_{stem}+CP+TAM$) which are individually extracted from spoken sentences. The verb phrases of the first column are declarative, the verb phrases of the second column are emphasized, and the verb phrases of the third column are pre-emphasized [3]. All three verb stems of each row bear the same tone.

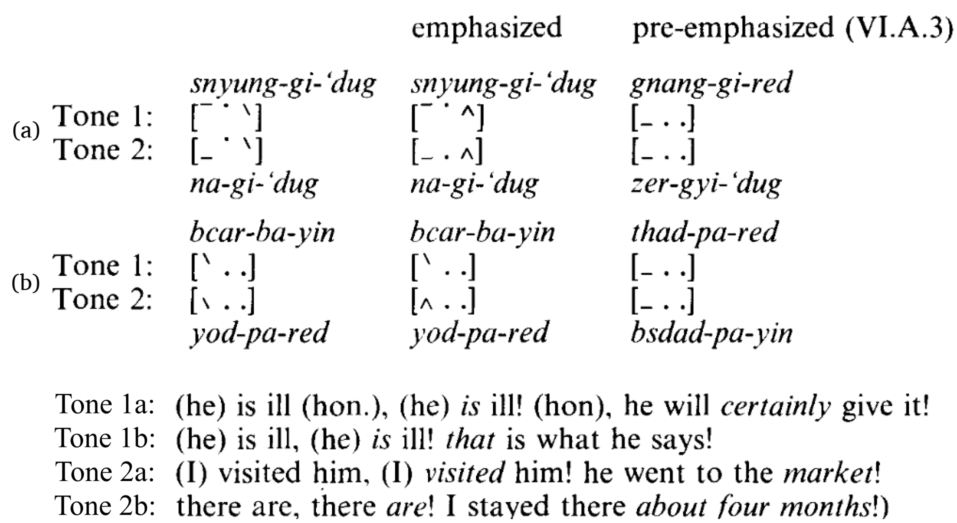


Figure 2.1: An example of Sprigg's (1993) junction analysis.

First of all, both Tone 1 and Tone 2 are neutralized in a pre-emphasized position, which illustrates tones of LT interacting with the information structure. This can be attributed to pitch range compression in the post-focal domain (cf. Monich,

[3]In the pre-emphasized phrase, the pitch pattern of this word shows that there is emphasis affecting some word earlier in the sentence or clause" (Sprigg, 1993, p.76). This phrase can be described as a post-focal domain based on current terminology (cf. Sugahara, 2003).

2011; Sugahara, 2003). In set (a), Tone 1 is realized as a high-level pitch contour, and Tone 2 is realized as a low-level pitch contour in both non-emphasized and emphasized contexts. In set (b), Tone 1 is realized as a high-falling pitch contour in both non-emphasized and emphasized contexts. Tone 2 is realized as a low-falling pitch contour in a non-emphasized context, but as a low-rising pitch contour when it is emphasized. The falling pitch contour is attributed to the following connecting particles (-pa/-ba)[4]. This also indicates that the tonal status of the grammatical particle is not negligible and needs to be included in the word tonology of LT (cf. Lim, 2010a). Sprigg's approach illustrates that Tone 1 and Tone 2 have a fairly stable pitch register, but interact with the immediate context (i.e., the connecting particle) and the intonational pattern (i.e., information structure) of the sentence. In short, his findings from the standpoint of a more rigorous analysis agree with the standard two-tone analysis in the respect that there are two contrastive lexical tones.

2.1.2 Four-tone analyses

The four-tone analysis is currently the leading analysis among Tibetan linguists in China. In this dissertation, however, the additional two tones are considered to be duration-conditioned allotones. Furthermore, the four-tone analysis, does not agree with the intuition of native speakers, who normally distinguish only high and low tones (Kong, 1995). For native speakers, the register of tones, high or low, is necessary for communication, but the contour of the tones is not (Tournadre and Dorje, 2003, p.36).

According to Hu (2002, p.200), Lhasa Tibetan language instruction at the Chinese Central Institute of Minority Languages has relied upon a four-tone analysis since the 1950s. In fact, Hu (1980) classified three types of four-tone system analy-

[4] /-pa/ ⟶ [-ba]/V_V

sis: Type 4A (Central University of Nationalities), Type 4B (Social Science Institute of China), and Type 4C (Chang and Chang, 1964). These three types have different notations but are otherwise virtually identical. These analyses recognize four citation tones: high level, high falling, low level, and low rising (Hu, 1980, 2002). This analysis is also accepted by some prominent linguists and Tibetologists outside China, such as Goldstein and Nornang (1970), and Kitamura (1974). The six-tone analysis, a direct variant of the four-tone analysis, also has some followers (Dawson, 1980; Hu, 1980). Table 2.3 compares the four-tone analysis with other mainstream analyses (Hu, 1980, p.28).

		Tone analyses				
WT	Gloss	2	6	4A	4B	4C
bka	‘order’	ka^1	ka^1	ka˥	ka˥˦	qā
ka.ba	‘pillar’	$ka\!:^1$	ka^2	kaa˥	ka˥	qāā
bkag	‘against’	$kaʔ^1$	ka^3	ka˥˨	kaʔ˥˦	qāà
sga	‘saddle’	ka^2	ka^4	ka˩˨	ka˦˥	qa̱
bsgar	‘install’	$ka\!:^2$	ka^5	kaa˩˧	ka˩˧	qa̱a̱
’gags	‘plugged’	$kaʔ^2$	ka^6	ka˨˩˨	kaʔ˧˥	qa̱à

Table 2.3: Comparison of four-tone analyses with other common analyses.

Although Hu (1980) identified six different types of pitch contour according to the segmental characteristics of the rhyme[5], he proposed a concessional four-tone analysis instead of the simpler two-tone or the more complicated six-tone analyses. In this approach, he imported segmental elements, such as vowel duration and glottal stop, into a suprasegmental or tonal classification. He recognized the strength of each subtype of four-tone analyses (4A, 4B and 4C), and did not have a preferred choice. However, he acknowledged that this analysis could be further reduced into two tonal categories: Tone 1 and Tone 2. Tone 1 is the lexical tone

[5] A similar approach is adopted by Fang Hu (2010). He introduces another pair of pitch patterns based on the aspirated onset and proposes eight tonal pitch contours. Again, he still has two major contrastive tonal categories: high and low.

that represents all the high onset allotones [55, 54, 52/53], whereas Tone 2 represents all the low onset allotones [12, 113, 132].

In conclusion, an analysis with four or more tones is not as phonologically economical and accurate as a two-tone analysis. Thus, a four-tone analysis could be more accurately described as an analysis with two lexical tones with four allotones. The same can be said of most six or eight tone analyses.

2.1.3 The tonal analysis of polysyllabic words

In this section, I will describe two main approaches to explaining the pitch pattern of polysyllabic words: the study of the so-called tone sandhi process between lexical tones, and the incorporation of the prosodic structure into the study of tonal patterns. The first approach holds the premise that the surface realization of a given word is the result of tone sandhi between the lexical tones of individual syllables or morphemes (Qu, 1981; Hu, 1980; Hu et al., 1982; Duanmu, 1992). The second approach, proposed by Gong (2014), integrates prosodic word structure into the analysis of the tone patterns of polysyllabic words. The prosodic domain defines the sandhi domain of the polysyllabic words.

2.1.3.1 Tone Sandhi of polysllyabic words

Qu provided a detailed description of the pitch pattern of disyllabic, trisyllabic, and quadrisyllabic words. The pitch pattern of the latter two is based on the pitch pattern of disyllabic words. According to Qu, the pitch pattern is the result of the tone sandhi of four underlying tones (/53, 55, 12, 14/). For disyllabic words, there are sixteen possible combinations for tone sandhi. I have reproduced the paradigm in Table 2.4, as it was later adopted by Duanmu (1992, p.10) in an influential autosegmental analysis that I will review in detail. I will start with Duanmu's analysis of the pitch pattern of disyllabic words and then proceed to the pitch pattern of trisyllabic and quadrisyllabic words.

No.	Rhyme length	Tone Sandhi	WT	Segments	Gloss
1.	s.s	/53.53/→[55.53]	*shug pa*	[ɕuʔ.pa]	'cypress'
2.	s.s	/53.12/→[55.53]	*mna' ma*	[na.ma]	'bride'
3.	s.l	/53.55/→[55.55]	*chu shel*	[tɕʰu.ɕel]	'crystal'
4.	s.l	/53.14/→[55.55]	*ha yang*	[ha.jaŋ]	'aluminium'
5.	s.s	/12.53/→[11.53]	*bdag po*	[taʔ.po]	'master'
6.	s.s	/12.12/→[11.53]	*tho rdo*	[to.ro]	'pile of rocks'
7.	s.l	/12.55/→[11.14]	*me phor*	[me.por]	'fire pan'
8.	s.l	/12.14/→[11.14]	*ja ril*	[tɕa.ril]	'bowl shaped tea brick'
9.	l.s	/55.53/→[55.53]	*khang pa*	[kʰaŋ.pa]	'house'
10.	l.s	/55.12/→[55.53]	*sla nga*	[laŋ.ŋa]	'earthenware pot'
11.	l.l	/55.55/→[55.55]	*bsam tshul*	[sam.tɕar]	'opinion'
12.	l.l	/55.14/→[55.55]	*shing sdong*	[ɕiŋ.toŋ]	'tree'
13.	l.s	/14.53/→[11.53]	*zam pa*	[sam.pa]	'bridge'
14.	l.s	/14.12/→[11.53]	*gdung ma*	[tuŋ.ma]	'roof beam'
15.	l.s	/14.55/→[11.14]	*ngal rtsol*	[ŋal.tsø]	'labor'
16.	l.l	/14.14/→[11.14]	*nyams myong*	[ɲam.ɲum]	'experience'

Table 2.4: Disyllabic pitch pattern of LT (Qu and Tan, 1983; Duanmu, 1992).

Duanmu proposes that surface forms can be derived based on two underlying tones: H and LH. Duanmu categorizes the pitch contours [55] and [53] as underlyingly H. Similarly, [12] and [14] are underlyingly LH. The variation of surface pitch contour of a given underlying tone or lexical tone is attributed to rhyme structure. Both [55] and [14] occur with longer sonorous rhymes (e.g. Vː, VN, etc.), while [53] and [12] occur with only simple or checked rhymes (e.g. V, Vʔ, etc.). In Table 2.4, the long rhyme is denoted by 'l', and the short rhyme is denoted by 's'.

As shown in Table 2.4, out of sixteen combinations of tone sets, there are only four possible sandhi form pitch patterns: [55.55], [55.53], [11.53], and [11.14]. The surface realization of a given disyllabic word depends on the underlying tone of the first syllable and the length of the second syllable, as shown in Duanmu's examples, repeated below:

(2) a. *shug pa* [ɕuʔ.pa] 'cypress.citation.'
53.55 → 55.55
/H.H/ → H.H
b. *ha yang* [ha.jan] 'aluminium (very.light)'
53.14 → 55.53
/H.LH/ → H.H
c. *bdag po* [taʔ.po] 'master.noun'
12.53 → 11.53
/LH.H/ → L.H
d. *nyams myong* [ɲam.ɲuŋ] 'roof.beam'
14.14 → 11.14
/LH.LH/ → L.LH

According to Duanmu (1992), the tone sandhi process of disyllabic words is governed by five basic syllable-based tone rules[6]:

(3) S0: Delete tones from non-initial syllables.
S1: Associate tones to syllables one-to-one, from left to right.
S2: If there are more syllables, spread the last tone to the excess syllables.
S3: If there are more tones, link the excess tones to the last syllable.
S4: If an L precedes a final long syllable with an H, spread L to the latter.

I show how these rules apply to (2c), *bdag po* [taʔpo], and (2d), *nyams myong* [ɲam ɲuŋ], to derive the pitch contour of both a disyllabic word that has an obstruent coda on the second syllable, and a word that has a sonorant coda, respectively:

(4) a. *bdag po* [tàʔ.pó] 'master':
S0: LH.LH→LH.∅
S3: LH.∅→L.H
b. *nyams myong* [ɲàm.ɲùŋ] 'experience':
S0: LH.LH→LH.∅
S3: LH.∅→L.H
S4: L.H →L.LH

[6]I have assigned an "S" prefix to the rules to remind the reader that they are syllable-based.

Qu and Tan (1983, p.36) claim that syllable grouping defines the tonal domains within a polysyllabic word. In other words, a polysyllabic word can consist of one or multiple syllable groups. All possible syllable groupings and corresponding surface tones of polysyllabic words in LT are presented in Table 2.5. Qu and Tan are not very explicit about the factors that define tone groups, but they are essentially of a semantic nature, as we will see in Chapter 3.

Syllabic type	Syllable grouping	Surface tones
Disyllabic	(σ.σ)	[H.H]
		[L.H]
Trisyllabic	(σ.σ.σ)	[H.H.H]
		[L.H.H]
	(σ.σ)(σ)	[H.H.H]
		[H.H.L]
		[L.H.H]
		[L.H.L]
	(σ)(σ.σ)	[H.H.H]
		[L.H.H]
		[H.L.H]
		[L.L.H]
Quadrisyllabic	(σ.σ.σ.σ)	[H.H.H.H]
		[H.H.L.H]
		[L.H.H.H]
		[L.H.L.H]
	(σ.σ).(σ.σ)	[H.H.H.H]
		[H.H.L.H]
		[L.H.H.H]
		[L.H.L.H]

Table 2.5: Syllable groupings and surfaces tones of polysyllabic words in LT.

According to Qu and Tan, the surface tones of the monosyllabic grouping (σ) remain the same as the underlying tone which can be either L or H. The surface tones of the disyllabic grouping (σ.σ) can be either L.H or H.H. There are also 'tonal templates' for polysyllabic words without any syllable groupings, which are (σ.σ.σ) and (σ.σ.σ.σ), and follow a series of predictable patterns. First, the surface tone of the word-initial syllable remains the same as the lexical tone, which is either L or H. Second, the surface tone of the second syllable is always H. Third, the surface tone of the word-final syllable for the trisyllabic grouping is always H. Lastly, the third and fourth syllables of the quadrisyllabic grouping follow the tone sandhi process of disyllabic grouping, which can be either L.H or H.H.

In short, the tonal domain is defined by the syllable grouping, and all the surface tones of polysyllabic words are mainly based on the tone sandhi pitch pattern of disyllabic words. That said, Qu and Tan (1983) and Duanmu (1992) are not able to account for all of the pitch patterns observed in this dissertation (see Chapter 2). For example, their 'templates' are not able to predict the location of downstep within a polysyllabic word (i.e., predictions of [H.H.H] and [H.H.H.H] are inaccurate). Furthermore, the syllabic groupings on which their analyses crucially hinge are also largely stipulated and are not derived by clear semantic and morphological factors.

2.1.4 Prosodic structure of polysyllabic words

The role of prosodic word boundaries in LT tone sandhi is not explicitly recognized in the earlier analyses (Sprigg, 1955; Hari, 1977; Qu, 1981; Hu, 1980; Dawson, 1980; Duanmu, 1990; Sun, 1997; Lim, 2012). In recent analyses of LT tonal and intonational phonology, however, it has taken a more prominent role (Monich, 2011; Gong, 2014; Lim, 2014, 2016), which is motivated by the fact that not all pitch patterns of LT polysyllabic words (e.g., downstep, the pitch pattern of affixes, etc.) are predicted by the earlier proposals mentioned above. Both Monich's work and my analysis will be presented in Chapter 3. For what remains of this subsection, I will focus on Gong's analysis.

Gong (2014) recognizes that prosodic word boundaries are essential to understanding the pitch pattern of morphologically complex words in Tibetan languages. Before proceeding to the discussion of the prosodic words, it is necessary to introduce Gong's non-mainstream approach to the categorization of LT tone. Gong classifies LT tones into two categories: register tones and contour tones. Monomoraic monosyllabic words either have a high-register (^{H}) or a low-register tone (^{L}). Bimoraic monosyllabic words also have a contour with additional contrast between level ($\mu\mu$) and falling ($\mu\grave{\mu}$). In other words, Gong considers that the pitch contour in LT is contrastive. Table 2.6 shows the possible tonal contrasts in LT monosyllabic words following Gong's classification (2014, p.3).

Register	μ	$\mu\mu$	
High	H [^{H}lo] *blo* 'mind'	HH [^{H}lɔɔ] *glo-ba* 'lung'	HL [^{H}lɔɔ̀] *klogs* 'Read!'
Low	L [^{L}lo] *lo* 'year'	LH [^{L}lɔr] *lor* 'paper money'	LML [^{L}lɔɔ̀] *logs* 'side'

Table 2.6: Possible tonal contrasts in LT monosyllabic words.

Interestingly, Gong claims that the appearance of a high tone in the low register tone (e.g. [^{L}lɔr] *lor* 'paper money') is a 'phonetic detail' that should be abstracted away. He proposes that the high tone originates from second-syllable stress (*σσ́)[7] of Proto-Tibetan proposed by Caplow (2009). During the tonogenesis process, it developed into second-syllable high tone. Gong argues that this high tone is not only realized on disyllabic words but also on monosyllabic ones. Thus, he proposes an automatic rule of H tone assignment to the last mora for monosyllabic

[7]It is important to point out that this stress pattern is only relevant to the non-verb word. The stress pattern for the verb word is on the first syllable (*σ́σ).

and disyllabic words alike (cf. Sun, 1997, p.503).

Gong (2014) also claims that all words in Tibetan languages (LT included) have a prosodic word structure. The prosodic word has two components: a maximally disyllabic *prosodic core* and an optional *prosodic appendix*. Figure 2.2 illustrates the structure of LT prosodic words.

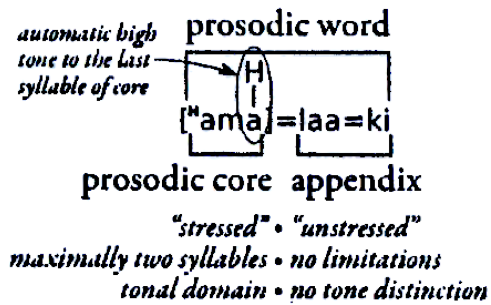


Figure 2.2: Structure of LT prosodic word (Gong, 2014, p.11).

The prosodic core is the tonal domain, and it is 'stressed' by receiving a high tone that is automatically assigned to the last syllable of the core. The prosodic appendix, however, is toneless and unstressed. All the syllables that are not included in the prosodic core are annexed to the prosodic appendix. Gong calls this a *push-out effect*. I have reproduced examples to illustrate the maximum disyllabicity and push-out effect in LT words in Table 2.7 (Gong, 2014, p.6).

σ	σ.σ	σ.σ.σ
[HH]	[HH.HH]	[HH.HL.LL]
[^{H}khãã]	[^{H}tsii.^{H}khãã]	[^{H}pẽẽ.^{L}tsøø̀].^{H}khãã
khang	*rtsis.khang*	*dpe.dzod.khang*
[house]	[count.house]	[book.treasury.house]
'house'	'account office'	'library'

Table 2.7: Maximum disyllabicity and push-out effect of LT words.

In this table, the word *khang* 'house' is compounded with a preceding monosyllabic word *rtsis* 'count' and forms the disyllabic word *rtsis.khang* 'account office'.

As the second syllable of the disyllabic compound, it is assigned a default high tone since it is the last syllable of the prosodic core. If the word *khang* 'house' is compounded with a preceding disyllabic word *dpe.dzod* 'book.treasury' to form the trisyllabic word *dpe.dzod.khang* 'library,' it is outside the maximally disyllabic prosodic core and in the prosodic appendix position. Therefore, the underlying H tone is not realized and the word receives a default low tone. I generally agree with Gong's application of prosodic boundaries (i.e., prosodic core), although I think invoking stress to account for the default H is unnecessarily complicated. However, I disagree with his dismissal of the role of the tones found at the prosodic appendix position, as this model is not able to account for all the pitch patterns discovered in this dissertation (see Chapter 3).

Gong also applies this analysis to suffixes, since affixation is one of the main morphological operations used to construct compound words. He claims that there are two types of suffixes: *tonic suffixes* (transcribed with a preceding '-'), and *atonic suffixes* (transcribed with a preceding '='). The tonic suffixes bear an H tone. It is realized as high-level pitch contour and neutralizes the falling pitch contour on the preceding syllable. Furthermore, tonic suffixes are included in the prosodic core if a position is available. The atonic suffixes, on the other hand, bear a L tone and are always assigned to the prosodic appendix. In Table 2.8 are examples of tonic and atonic suffixes (Gong, 2014, p.7).

stem-tonic suffix	stem=atonic suffix
phøð̀ 'escape' + *-ki* 'NPST'	*phøð̀* 'Tibet' + *=ki* 'GEN'
[LL.H]	[LML.L]
/LML-H/	/LML=∅/
[^{L}phyy.ki]	[^{L}phøð̀].ki
byol.ki	*bod.ki*

Table 2.8: Minimal pairs of tonic and atonic suffixes.

Thus, the contrast between tonic and atonic suffixes is only revealed when they are affixed to a monosyllabic word. The tonal status of suffixes is irrelevant in polysyllabic words. Again, I disagree with Gong's analysis of the tonal status of suffixes. I argue that all affixes are tonal (H or L) and play important roles in the word tonology of LT (see §3.2.1 and (Lim, 2010a)).

Gong's proposal is interesting, especially given his insights into prosodic structure for the analysis of polysyllabic words (cf. Lim, 2013, 2016). In fact, there are many parallels between Gong's proposal and the approach espoused in this dissertation. However, there is also a difference between the two analyses that warrants deeper evaluation (see Chapter 3).

2.2 Previous research on word stress

From Caplow's (2009) research, we understand that there was stress in Proto-Tibetan. Stress is preserved in the archaic non-tonal Tibetan languages, but there is still uncertainty about the existence or functional load of stress in the innovative tonal Tibetan languages. The presence of stress can inevitably complicate the analysis of tonal and intonational phonology in these languages.

There are claims that there is a stress-like system in LT, but these claims have not been sufficiently investigated to reach a consensus (Denwood, 1999; Odden, 1979; Tournadre and Dorje, 2003). Meredith (1990) has provided, by far, the most systematic phonological analysis of LT stress. Meredith proposes a *Stressability Hierarchy* in which stress is phonologically assigned to heavy syllables. If there are two equally heavy syllables in a word, tone determines stress placement: syllables with high tone are preferred. If there is still a tie, the leftmost syllable in a word is stressed. Table 2.9 provides some examples that illustrate the stress and tone assignment of LT based on Meredith's hierarchy of rules.

WT	IPA	Gloss
rda ram sa la	[dà.ˈrám.sà.là]	‘Dharamsala’
ka ra dkar po	[ká.rá.ˈkár.pó]	‘white sugar’
mi dge ba	[mì.ˈgé.wà]	‘non-virtue’
mi bzod pa	[mì.ˈɕɵ́ː.pà]	‘unbearable’

Table 2.9: Tone and stress assignment based on Meredith’s Stressability Hierarchy

Meredith’s analysis of stress in LT appears to correspond to a non-native perceived prominence since the stress assignment is associated with syllable weight, pitch, and syllable position, three factors that play a major role in English stress and cross-linguistically. Although stress is not the main research focus of this dissertation, I will present results from some of my preliminary instrumental investigations into LT stress in Chapter 3 and discuss their implication in analyzing the tonal and intonational phonology of LT.

Some analyses also invoke the notion of stress to explain the systematic presence of a H tone on the second syllable of disyllables (Monich, 2011; Gong, 2014). They essentially reanalyze the modern reflexes of what was most probably stress at an earlier stage of the language (Caplow, 2009) as a non-typical form of synchronic stress. In Chapters 3 and 4, I will argue that it is unnecessary to analyze this H tone as a stress marker.

2.3 Previous research on intonation

In this section, I will explore the traditional understanding and modern investigation of intonation in LT. There is relatively little research in this area. I will first provide a concise overview of prosodic structure, which plays an important role in the word tonology of LT. Then, I will present a brief discussion of Autosegmental Metrical Theory, since this is the main framework of intonational phonology used in this dissertation. Subsequently, I will discuss the work of Monich (2011), which is currently the only paper that focuses on LT intonational phonology. Lastly, I will briefly comment on a few other papers that focus on LT speech synthesis research.

2.3.1 Prosodic Structure

This dissertation focuses on the phonology of tone and intonation, but these are not autonomous, closed systems: they interact with segmental phonology and other non-phonological systems, such as morphology and syntax. In this section, I will briefly review the theoretical literature on prosodic domains defined by the morphosyntax. This is crucial because, as we will see in the rest of the thesis, prosodic constituency affects both LT lexical tone and intonational targets.

Researchers such as Selkirk (1980) and Nespor and Vogel (1982) have postulated that the interaction between phonology and morpho-syntax supports a hierarchical model of prosodic constituents. In their seminal work, Nespor and Vogel (1984, 2007) identify the following prosodic constituents: phonological utterance (ν), intonational phrase (ι), phonological phrase (ϕ), clitic group (C), phonological word (ω), foot (f), and syllable (σ). Later on, the mora (μ) was also inducted as one of the prosodic constituents (Hayes, 1989a; Ito and Mester, 2008; McCarthy and Prince, 1993; Pierrehumbert and Beckman, 1988). These constituents form a hierarchical structure, where larger constituents dominate the smaller constituents as shown in Figure 2.3.

There is a strong tendency to maintain the correspondence between syntactic and prosodic constituency cross-linguistically. This correspondence of syntactic-prosodic constituency is captured by a set of universal Match constraints outlined in Match Theory (Selkirk, 2011). These constraints require a constituent (Clause/Phrase/Word) of a syntactic structure be complemented by a corresponding constituent ($\iota/\phi/\omega$) of a prosodic structure. If the phonological markedness

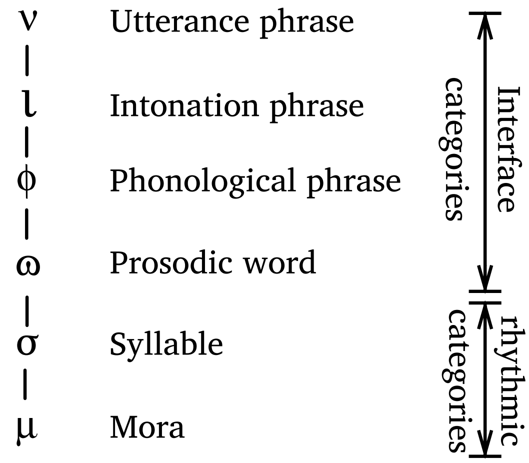


Figure 2.3: A prosodic hierarchy

constraints of a given language are higher ranked than Match constraints, its prosodic structure violates the *Strict Layer Hypothesis* [8]. As a result, some languages exhibit characteristics of recursivity and level-skipping.

According to Ito and Mester (2008), these constituents can be further divided into two main classes: the interface categories and the rhythmic categories. The interface categories are comprised of syntactic constituents (phonological utterance, intonational phrase, phonological phrase, clitic group, phonological word). They are usually defined by the phonological interaction between morphosyntax and phonological constituents. The phonological utterance is the largest constituent in the prosodic hierarchy, and it is related to the syntactic sentence. This corresponds to the top node (X_n) that dominates the rest of the syntactic tree (Nespor and Vogel, 2007, p.222). This prosodic category type, however, is not posited in the prosodic hierarchy of LT in my own proposal. Intonational phrases are related

[8]"The strict layer hypothesis: A constituent of category-level n in the prosodic hierarchy immediately dominates only a (sequence of) constituents at category-level n-1 in the hierarchy" (Selkirk, 2011, p.437).

to syntactic clauses (Truckenbrodt, 2007, p.436). These include “parenthetical expressions, non-restrictive relative clauses, tag questions, vocatives, expletives, and certain moved elements” (Nespor and Vogel, 2007, p.188). Phonological phrases are related to syntactic phrases. Typical syntactic phrases include noun phrases, verb phrases, and adjective phrases. Nespor and Vogel also recognize the clitic group as an independent prosodic domain, which is, by far, the most controversial constituent in the prosodic hierarchy (Vigário, 2010). This is because clitics are usually either associated with a phonological phrase as an independent word, or affixed to a phonological word (Nespor and Vogel, 2007, p.145). Phonological words (also known as Prosodic words) are usually considered as constituents that occupy the terminal node of a syntactic tree. They are syntactically related to individual stems, stems of a compound word, and affixes (Nespor and Vogel, 2007, p.110).

The rhythmic categories are comprised of sub-prosodic word constituents (foot, syllable, mora), and refer to morphology. They are defined by their intrinsic phonetic characteristics and prominence patterns. The foot is the domain of stress and normally groups stressed and unstressed syllables (Hayes, 1976). The syllable is probably the most common prosodic unit crosslinguistically. It is a common tone and accent bearing unit across languages. The mora is the smallest prosodic constituent and has no direct morphosyntactic counterpart. It is usually used to phonologically represent the weight or duration of a syllable. It is also a tone or accent bearing unit in some languages.

LT appears to follow the strict layer hypothesis (Selkirk, 1984; Nespor and Vogel, 1984). In spontaneous speech, it shows no evidence of prosodic recursion or of non-exhaustive parsing. However, it does not seem to strictly abide by the universal prosodic hierarchy: I will argue that it does not have a phonological phrase, but requires an additional level above the prosodic word (Vigário, 2010). I propose that it is a *Prosodic Word Group* (PWG), an independent domain, distinct from the clitic group[9] and located below the phonological phrase, that consists of several prosodic words. According to Vigário (2010, p.498), the main evidence to support the prosodic status of the prosodic word group is shown by the following: "*a*) the perception/intuition of word-level stress; *b*) segmental phenomena cueing stressed and stressless vowels, such as the absence or presence of vowel reduction and semi vocalization; *c*) segmental and suprasegmental phenomena cueing the right edge and the left edge of the prosodic word; *d*) pitch accent distribution; and *e*) deletion processes referring to the prosodic word, namely clipping and deletion under identity, among others." I will show that in LT, a downstep phenomenon applying at the right and left edges of prosodic words suggest the existence of a prosodic word group. In addition, the boundary of the prosodic word group coincides with the word boundary of the polysyllabic word. A detailed discussion of the role of prosodic structure and the tonal/intonational phonology of LT will be presented in Chapter 3 and Chapter 4.

2.3.2 Autosegmental Metrical (AM) Theory

Autosegmental Metrical (AM) models were developed as a merger of autosegmental phonology and metrical models (Bruce, 1977; Goldsmith, 1976; Pierrehumbert, 1980; Pierrehumbert and Beckman, 1988; Liberman, 1975; Ladd, 1996). Ladd (1996, p.10-14) provides an insightful commentary on the development of intonational phonology. He distinguishes two approaches to intonational research in earlier studies: impressionistic or proto-phonological, and instrumental or phonetic. While these two camps had different theoretical approaches, they are not neces-

[9] It is one of the most contentious constituents in original models of the Prosodic Hierarchy (cf. Hayes, 1989b; Vigário, 2003, 2010).

sarily irreconcilable. Ladd proposes that modern intonational phonology should combine the tasks of identifying the "categorically distinct entities or phonological constituent" and establishing the relationship between phonetic realization and corresponding phonological entities (1996, p.11).

Ladd (1996, p.42-43) identifies four major characteristics of AM:

(5) a. *Linearity of tonal structure*
The pitch contour of an utterance consists of a sequence specified *local event*, associated with specific points of the segmental stream. The connection between the local events is termed as *transition* and it is phonologically unspecified. The most important local event is pitch accents, which are often associated with prominent segments. The edge tones or boundary tones indicate the onset and offset of an utterance.

b. *Distinction between pitch accent and stress*
First and foremost, pitch accent is an *intonational* feature associated with specific syllables according to the intonational structure of a given language. In some languages, it is also used to characterize prominence for certain syllables, which is perceived as stress. If the stress is metrical, then it can be differentiated from pitch accent.

c. *Analysis of pitch accents regarding level tones*
All pitch accents are composed of primitive level tones, which can be analyzed as high or low.

d. *Local sources for global trends*
The global pitch contour is the summation of the F0 effect contributed by local tone and accent. Thus there is no superposition of pitch contour (i.e. global pitch contour + local pitch effect). Depending on the position of the pitch contour, there is a scaling factor for the local tone and accent.

Based on research on English and other Indo-European languages, the proposed hierarchical structure of intonational tones consists of: *a*) boundary tone (initial & final); *b*) pre-nuclear accents; *c*) nuclear accent; and *d*) phrase tone (Pierrehumbert, 1980; Ladd, 1996). In the following, I will provide an overview of the characteristics of these intonational tones.

In the AM framework, intonational tones have only two contrastive privative level tones—High (H) and Low (L) tones (Pierrehumbert, 1980; Beckman and

Pierehumbert, 1986; Ladd, 1996). They can take two forms. Tones can associated with the boundaries of prosodic domains (edge tones) or can be aligned with the stressed syllable (pitch-accents). In the AM framework, there are two types of pitch accents: mono-tonal (i.e., L*, H*) or bitonal (i.e., L* + H, L + H*, H* + L, H + L*)[10]. Thus, there were only six types of pitch accent contours in the inventory (Beckman and Pierehumbert, 1986, p.256). Intonational languages often utilize pitch accents as the primary way of marking information structure of a given utterance. For instance, the H* + L falling tone is a common cross-linguistic strategy to mark focus, even if it can be accompanied by other focus-marking strategies (Hartmann, 2007, p.225), such as a higher pitch excursion and a post-focal deaccenting. The focused constituent has a higher pitch excursion, but the post-focal pitch contour is often deaccented.

Edge tones can be divided into boundary and phrasal tones. Boundary tones are usually a single H or L mono-tone assigned to the onset or offset edges of an intonation phrase or utterance phrase. They are usually designated by a '%' symbol. If the boundary tone is in the onset phrase position, it usually precedes the tone (e.g.,e.g., %H), but it follows the tone if the tone is in the phrase offset position (e.g., L%). A phrase-final H% is often associated with a question and conveys continuation, while a phrase-final L% tone is often associated with a statement and conveys finality (Gussenhoven, 2002, p.95). Phrasal tones or phrase accents are marked with '$^-$' or unmarked. They are intervening tones between the preceding pitch accent and a boundary tone. According to Beckman and Pierehumbert (1986), they are the edge tones of an intermediate interface prosodic constituent, such as the phonological phrase.

[10]Bitonal accent is also known as accent contour. The H + H* bitonal pitch accent found in the original version is eliminated because it is indistinguishable from other sequences due to neutralization (Pierrehumbert, 1980).

I give a schematic representation of the structure of tunes in Figure 2.4. With two types of boundary tones, six-pitch accents, and two phrase tones, it is possible to generate twenty-two combinations that parallel the British-style nuclear tones description (Ladd, 1996, p.82).

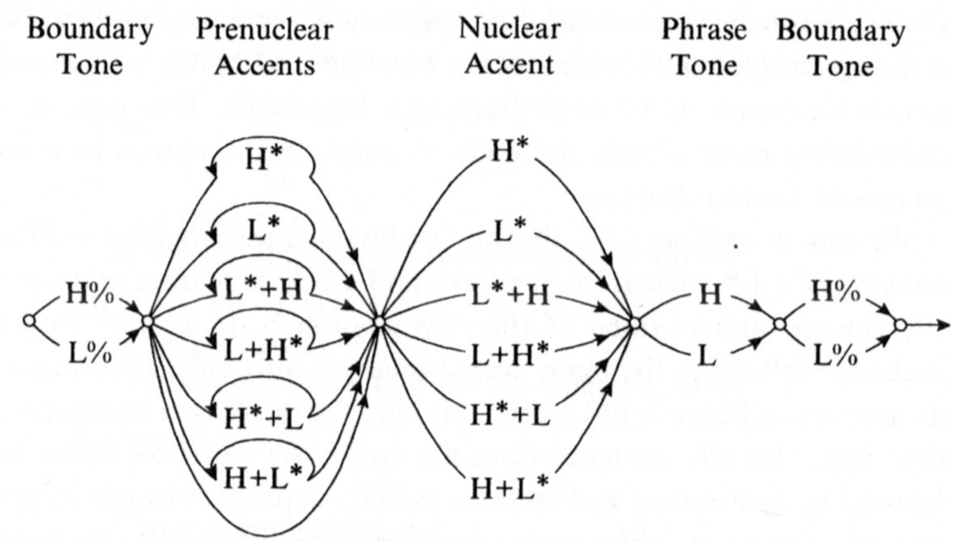


Figure 2.4: Finite-state grammar of generating tunes (Ladd, 1996, p.82).

In the AM framework, both word tones and boundary tones are treated as phonetic pitch targets, and the intonation contour of an utterance is constructed through the interpolation between targets. The phonetic effect of boundary tones are more localized, and they tend to modulate the pitch contour at the periphery of prosodic constituents. Yip (2002, p.260-261) proposed that there are four mechanisms that an intonational tone contributes to the overall pitch contour of a phrase.

(6)
a. The entire pitch register of the prosodic constituents is shifted higher or lower, but lexical tones still maintain their relative pitch contrast.
b. The pitch range of the prosodic constituent is expanded or contracted, where the relative pitch contrast has increased or decreased.
c. The boundary tones at the edge of prosodic constituent may attach to the closest tone bearing unit during pitch realization.

d. The register shifting phenomenon can occur within an interface prosodic constituent but it can be suspended or reset at the prosodic boundary.

These mechanisms are more straightforward and readily observed in non-tonal languages, but in languages such as LT that have lexical tone, things are far more complex. The interplay between the lexical and intonational tones of LT will be further investigated in Chapter 4.

2.3.3 Previous work on the interaction between tone and intonation in LT

Monich (2011)

Monich (2011) provides important insights into the interaction between tones and intonation in LT. She argues that LT has a trochaic metrical system (see §3.1.2), and that the intonation of LT is based on prosodic phrasing: feet, prosodic words, and accentual phrases. Monich (2011, p.16) proposes that a LT morpheme has a lexical tone (L_{lex}or H_{lex}), followed by a H* tone which diachronically corresponds with the pitch peak of a stressed syllable. Thus, the foot of a prosodic word can bear one of the two underlying bi-tonal melodies: ($L_{lex}+H$*) or ($H_{lex}+H$*). However, LT grammatical particles (e.g., connecting particles, case maker, etc.) do not bear lexical tone in her analysis, and they either form their own feet or are included with the preceding feet. The tones of the strong syllable (σ_s) of a foot are preserved, while the lexical tones of the weak syllable (σ_w) are deleted and replaced by the preceding H-tone (H*). Subsequently, the tonal contour of words is further adjusted when they are organized into accentual phrases. LT information structure also interacts with prosodic structure, in that the pitch accent of the focused constituent in each LT accentual phrase is preserved (usually at phrase-initial position). Monich claims that all the lexical tones of all non-focused syllables are de-emphasized post-lexically.

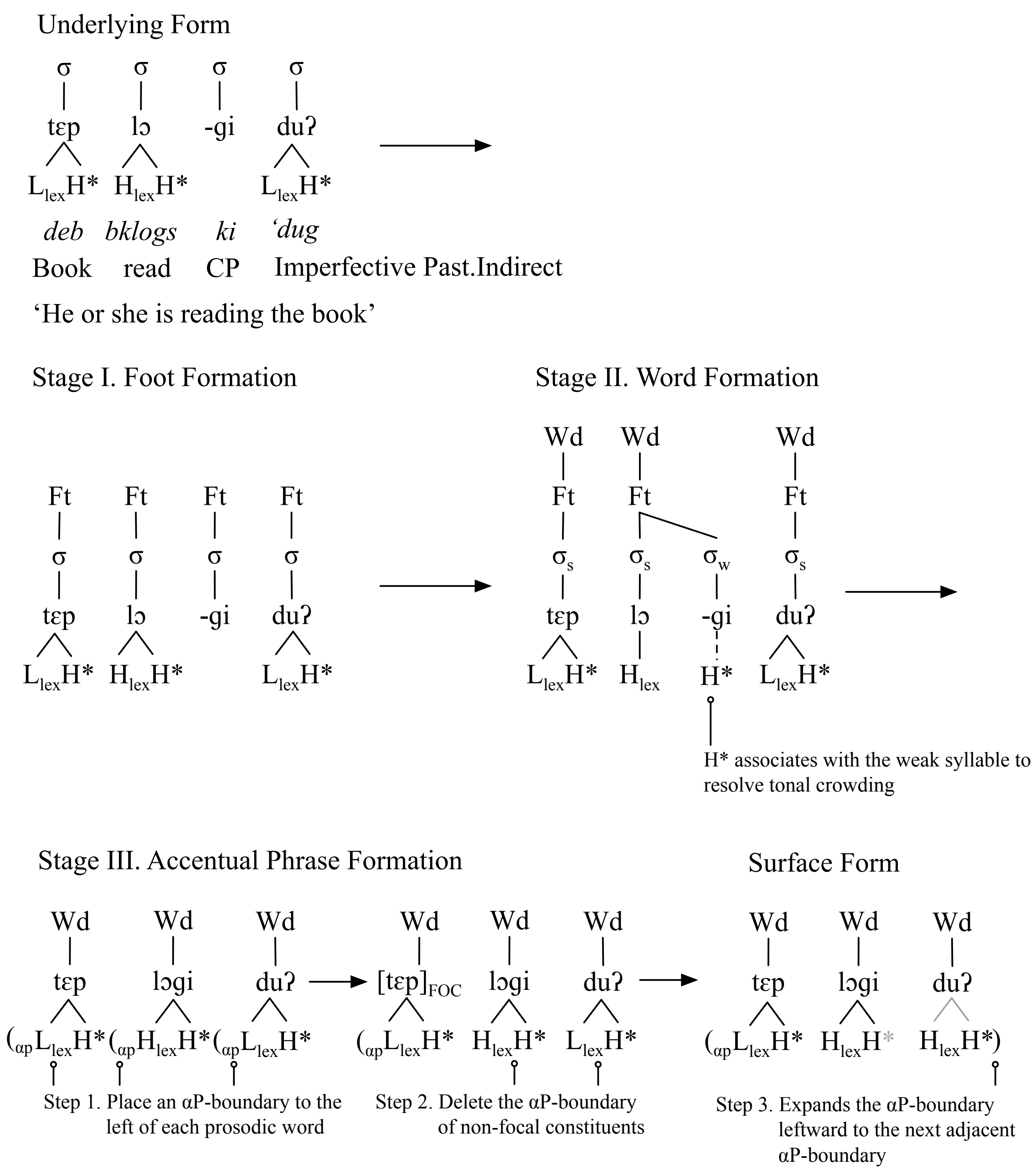


Figure 2.5: The derivation of the surface intonational tone of LT (adapted from Monich, 2011, p.17).

I illustrate Monich's principles of LT intonation formation with the *deb bklogs ki 'dug* [tɛp lɔ.gi duʔ] 'He's reading a book,' shown in Figure 2.5.

At stage I, a foot is assigned to each morpheme in the utterance, i.e. [((tɛ̀p)$_{\sigma}$)$_{Ft}$((lɔ́)$_{\sigma}$)$_{Ft}$((gì)$_{\sigma}$)$_{Ft}$((dùʔ)$_{\sigma}$)$_{Ft}$].

Stage II is the *formation of prosodic words*. According to Monich, three prosodic words are formed, where the first and third words are monosyllabic prosodic words with a degenerate foot (i.e. (tɛ̀p)$_{Wd}$ and (dùʔ)$_{Wd}$), but the word composed of a noun and a suffix gets merged into a single foot ((lɔ́)$_{\sigma}$)$_{Ft}$((gì)$_{\sigma}$)$_{Ft}$$\rightarrow$ (((lɔ́)$_{\sigma_s}$(gí)$_{\sigma_w}$)$_{Ft}$)$_{Wd}$). Since LT has a trochaic metrical system, the initial syllable of this disyllabic foot is a strong syllable ((lɔ́)$_{\sigma_s}$), and the final syllable is a weak syllable ((gì)$_{\sigma_w}$). The H* is initially assigned to the prominent strong syllable, but since there are two tones on a syllable with a short rhyme, the preceding H* is assigned to the weak syllable to resolve tonal crowding. In Duanmu's (1992) model, this is achieved by the deletion of the non-initial tone.

Stage III is the *formation of accentual phrases (αP)* and follows the prescribed steps below (Monich, 2011, p.39):

1. Assign a left αP-boundary to the left edge of each prosodic word.
2. Remove the informationally insignificant (i.e., non-focal constituents) αP-boundaries unless they are at the onset position of an intonation phrase (iP).
3. Extend leftward each left αP-boundary to the adjacent αP-boundary or the offset of the iP to form an αP-domain.

In the current example, only the phrase-initial prosodic *deb* [tɛp] 'book' is in focus. According to Monich's observations, the pitch range of the leftmost H* in individual focal and pre-focal constituents is expanded, but there is neither compression nor modulation of the pitch range post-focally.

In short, Monich offers a sophisticated approach to the account of LT tonal and intonation patterns. That said, I have some reservations concerning her approach, in particular, the reality of the metrically based stress system. Besides, the absence of boundary tones in her version of intonational phonology is incompatible with the results obtained from my data. I will provide my assessment on the tonal and intonational phonology of LT in Chapter 3 and Chapter 4, respectively.

2.3.4 LT Speech synthesis

In the current dissertation, I propose a synthesis model of LT intonation as a way of confirming the validity of my analyses on the tonal and intonational phonology of LT. Currently, there are to my knowledge only two papers that specifically focus on the speech synthesis of LT: Guo et al. (2011) and Chen Qi (2010). Both researchers used predefined speech prosodic parameters and a respiratory signals model (expansion/contraction of chest and abdomen circumference) to synthesize the pitch contours of LT. They focused on modelling the duration and amplitude of the 'breathing signals'. Based on this data, the researchers then developed an Radial Basis Function (RBF) neural network to predict the prosodic pattern. In other words, they are a non-linguistic, engineering-based approach to speech synthesis.

It is important to point out that neither of these models were verified by rigorous statistical analyses, which is why I need to rely on their reports as well as my personal judgement. The synthesized speech by Chen Qi (2010) was evaluated using a subjective test of perception and recognition based on a scale from 1 (unacceptable) to 5 (natural). The highest reported result is 3.37 which is considered to be acceptable. From a cursory look, the synthesized pitch contour looks like a reasonable approximation of the original pitch contour. Guo et al. (2011) show that the duration and frequency difference between the target and the synthesized speech is within 60 ms and 40 Hz respectively. According to them, the results are considered good, but there is no way to verify this. I am concerned about the frequency difference, where a difference of 40 Hz might mask some crucial linguistic details such as downstep. In short, both results produced reasonably good synthesized intonation patterns using sophisticated technology and algorithms. In Chapter 5, however, I will present my pitch synthesis model based on tonal and intonational phonology, which has the ability to produce equally good or better synthesized pitch contours.

Chapter 3

The word tonology of LT

The aim of this chapter is to present a description of the word tonology of LT. In §3.1, I argue that word stress plays no role in the word tonology of LT. In §3.2, I address the tonal status of affixes (§3.2.1). Subsequently, I establish that the *prosodic structure* (§3.2.2) and the *Revised LhaSa tone rules (RLS)* (§3.2.3) are the two main components of LT word tonology. I also present the results of instrumental investigations concerning the surface realization of the tones of LT words (§3.3) and the status of word stress in LT words (§3.4).

3.1 Word stress in LT

There is no consensus concerning the existence and characteristics of Tibetan word stress. One of the main reasons is that it is often difficult to differentiate between the formal representation and the correlates of 'stress' (cf. van der Hulst, 2012). To avoid unnecessary vagueness, the definition of word stress has to be both phonologically and phonetically meaningful. Phonologically, word stress is a manifestation of word-level metrical structure (Hayes, 1976, p.8). According to Hyman (2006, p.231), it structure needs to comply with the following requirements:

1. *obligatoriness*[...]: every lexical word has at least one syllable marked for the highest degree of metrical prominence (primary stress);
2. *culminativity*[...]: every lexical word has at most one syllable marked for the highest degree of metrical prominence.

Furthermore, it is crucial to differentiate word stress from phrasal stress. Common phenomena that are attributed to phrasal prominence are phrase/word-final lengthening and pitch modulation caused by boundary tones and information structure (Brunelle, 2017, p.2). Researchers might not able to isolate the effect of phrasal stress if the data is only collected in citation form or is not framed properly in a carrier sentence (cf. Ladefoged, 2003, p.7-8). The failure of earlier LT research to distinguish between word and phrasal prominence may have had an effect on previous analysis and findings (see Chapter 4).

Phonetically, word stress is usually cued by an increase in syllable duration, intensity, and pitch (Hayes, 1995; Fletcher, 2013). Changes in vowel quality (e.g., loss of phonological contrast) in unstressed syllable is another important cue (Hayes, 1976, p.84). A problem often faced by researchers working on tone languages is that pitch can be a phonetic cue of both stress and tone. This does not necessary exclude the coexistence of lexical tone and word stress, as in Swedish, Ma‘ya, and Mandarin (Riad, 2013; Remijsen, 2002; Duanmu, 2002), but it can greatly complicate the analysis. The phonetic correlates of word stress will be investigated experimentally in the later part of this Chapter (§3.4). In this section, I will review the phonological evidence for stress in LT and try to disentangle the often contradictory accounts that have been previously proposed.

The two main supporters of the role of word stress in LT tonal phonology are Meredith (1990) and, more recently, Monich (2011). Both claim that a stress system governs the neutralization and surface realization of tones, but the stress systems that they describe are dramatically different.

3.1.1 Meredith (1990)

Meredith (1990, p.42-126) claims that the metrical head of a word is determined by the morphological structure of the word, and by the "perceived stressability" of syllables that is dependent on syllable weight and underlying tone (cf. §2.2). The perceived stressability is organized into the following hierarchy: Heavy syllable/High register > Heavy syllable/Low register > Light syllable/Hight register > Light syllable/Low register (Meredith, 1990, p.86). In general, the rightmost syllable that bears the H tone in a disyllabic word is the metrical head (Meredith, 1990, p.70). Only the tone of the metrical head is kept, and the other syllables receive an L by default.

According to Meredith's analysis, LT appears to be a right-dominant, weight-sensitive, unbounded and right-to-left stress system (cf. Hayes, 1976). However, it appears that this stressability hierarchy is, just a form of perceived prominence and is not phonologically motivated. When one looks at the data carefully, tonal assignment is independent of syllable weight, and the tones of "unmetrified" syllables are still present on the surface. The fact that Meredith's description is so dfferent from all other accounts may be attributed to the fact that his speaker's mother tongue is *kham.pa.tsang*, and that he acquired Standard Tibetan through the traditional Tibetan schooling system, and by living among the Tibetan Diaspora in Nepal (Meredith, 1990, p.42). This is a concern, since the tonal pattern of his mother tongue *kham.pa.tsang* is an "intersection between tonal melody and voice register" (Thurgood and LaPolla, 2003, p.687). I will therefore assume that Meredith's data is not comparable to that discussed elsewhere.

3.1.2 Monich (2011)

Monich proposes that the stress system of proto-Tibetan has lost its substance, but that metrical structure has nonetheless remained relevant for vowel harmony and tonal feet (Monich, 2011, p.15). She argues that progressive vowel harmony (from left (stressed) to right (unstressed), and vowel reduction in historically unstressed syllables provide evidence to support a trochaic metrical system. In progressive vowel harmony, the vowel of the final syllable assimilates with vowel height of the preceding vowel, as shown in the examples reproduced in Table 3.1. For instance, there is an alternation of the vocalic nuclei [ø∼y] in example (a) of Table 3.1. If the vocalic nucleus of the first syllable is [a], the vocalic nucleus of the following syllable is simply [ø], as in [gʲàbǿː] 'Sino-Tibetan.' However, if the vocalic nucleus of the first syllable is [i], the vocalic nucleus of the following syllable is raised and fronted to [y] to agree with [i].

a.	*rgya.bod*	/gʲa$^{\text{LH}}$.bøː$^{\text{LH}}$/	[gʲà̱bǿ̱ː]	'Sino-Tibetan'
	dbyin.bod	/ʔin$^{\text{H}}$.bøː$^{\text{LH}}$/	[ʔí̱mbý̱ː]	'Anglo-Tibetan'
b.	*rgya.skad*	/gʲa$^{\text{LH}}$.kɛː/	[gʲà̱gɛ̱́ː]	'Chinese language'
	dbyin.skad	/ʔin$^{\text{LH}}$.kɛː/	[ʔí̱ŋgé̱ː]	'English language'

Table 3.1: Examples of LT progressive vowel height harmony process (adapted from Denwood, 1999, p.79).

According to Denwood (1999, p.78), the vowels [a, ɛ, ɔ] in short open initial syllables or monosyllables are reduced to [ə, ɪ, o/ʊ] in non-initial syllables. In monosyllabic adjective particles, the underlying vowels /a, o, e/ are also reduced to [ə, ʊ, ɪ] respectively. Examples of these processes are shown in Table 3.2. This pattern of vowel reduction seems to support a trochaic parse.

Upon closer examination, however, both pieces of evidence from are inadequate to support the existence of a LT trochaic system. First, the LT vowel harmony process is bidirectional, and this poses a significant challenge to the claim

a.	a ∼ ə	*lta.ba.*	/ta^{LH}.ba^{L}/	[tà̱wə́]	‘view’
		da.lta.	/da^{LH}.ta^{H}/	[tàndə̱́]	‘now’
b.	ɛ ∼ ɪ	*rtse.mo.*	/tse^{H}.mo^{LH}/	[tsɛ̱́mʊ́]	‘tip’
		rgyal.rtse	/gʲal^{LH}.tse^{H}/	[gʲàndzɪ̱́]	‘Gyantse’ (place)
c.	ɛ ∼ ɪ	*deb.*	/deb^{LH}/	[tɛ̱̀p]	‘book’
		phyag.deb.	/tçʰa^{H}.deb^{LH}/	[tçʰáːdɪ̱́p]	‘book’ (hon.)
d.	ɛ ∼ ɪ	*’phreng.ba*	/treŋ^{H}.ba^{LH}/	[trɛ̱́ŋə́]	‘rosary’
		phyag.’phreng	/tɕʰan^{H}.treŋ^{H}/	[tɕʰántrɪ̱́ː]	‘rosary’ (hon.)
e.	ɔ ∼ o	*mtsho.*	/tsʰo^{H}/	[tsʰɔ́]	‘lake’
		rgya.mtsho	/gʲa^{LH}.tsʰo^{H}/	[gʲà.dzó]	‘sea’
f.	ɔ ∼ o	*blo.*	/lo^{H}/	[lɔ́]	‘mind’
		bsam.blo.	/sam^{H}lo^{H}/	[sámló]	‘thought’

Table 3.2: Examples of apparent vowel reduction conditioned by stress alternations (adapted from Denwood, 1999, p.78).

that it is based on a trochaic parse. There are many examples which support the existence of regressive vowel harmony (Dawson, 1980, p.63-72). In fact, Dawson (1980, p.82) claimed that “progressive raising [vowel harmony] is less strong than regressive raising.” Thus, if there is a metrical structure in LT, it is the rightmost syllable of a disyllabic word that needs to be in the stronger position relative to the vowel harmony. The latter observation supports an iambic system, in agreement with Meredith (1990) and is consonant with the historical reconstruction of Proto-Tibetan stress that places it on the second syllable of disyllabic words (Caplow, 2009, p.545). I have reproduced some regressive vowel harmony examples based on Dawson’s original transcription in Table 3.3 and Table 3.4. In short, the LT stress pattern proposed by Monich does not match the direction of the vowel harmony process observed by Dawson (1980).

		Citation form			Past tense *pa red* /-pə^{L}.reː^{L}/ 's/he __ '	Future tense *ki red* /-qi^{H}.reː^{L}/ 's/he will __ '
a.	e ~ i	*bzhes*	/ʃeː^{LH}/	'get' (H)	[ʃèːpə̀.rèː]	[ʃìːqí.rèː]
b.	ɛ ~ ɪ	*zer*	/ʃɛː^{H}/	'say'	[ʃɛ́ːpə̀.rèː]	[ʃɪːqí.rèː]
c.	o ~ u	*btsongs*	/tsõː^{H}/	'sell'	tsṍːpə̀.rèː]	[tsṹːqí.rèː]
d.	ɔ ~ ʊ	*bklogs*	/lɔː^{H}/	'read'	lɔ́ːpə̀.rèː]	[lʊːqì.rèː]
e.	ø ~ y	*mchod*	/tʃʰøː^{H}/	'eat' (H)	[tʃʰǿːpə̀.rèː]	[tʃʰýːqí.rèː]
		Citation form			*ku*[1] /-ku^{H}/ 's/he has not yet __'	*dus* /-tyː^{H}/ 'when s/he __ '
a.	e ~ i	*bster*	/teː^{H}/	'give'	[tíːkú]	[tíːtýː]
b.	ɛ ~ ɪ	*bsnyal*	/ŋɛː^{LH}/	'sleep'	[ŋ̊ìːkú]	[ñìːtýː]
c.	o ~ u	*btsongs*	/tsõː^{H}/	'sell'	[tsúːkú]	[tsúːtýː]
d.	ɔ ~ ʊ	*bklogs*	/lɔː^{H}/	'read'	[lʊ́ːkú]	[lʊ́ːtýː]
e.	ø ~ y	*bros*	/pʰøː^{LH}/	'flee'	[pʰỳːkú]	[pʰỳːtýː]

Table 3.3: Examples of regressive vowel height harmony in verb constructions (Dawson, 1980, p.64, 67).

		Citation form					Compound word	
a.	e ~ i	*me*	/me^{LH}/ 'fire'	+	*shing*	/ʃiŋ^{H}/ 'wood'	[mìʃíː]	'firewood, fuel'
b.	ɛ ~ ɪ	*bsnyal*	/ŋɛː^{LH}/ 'sleep'	+	*gur*	/qʰuː^{LH}/ 'tent'	[ŋ̊ìːqúː]	'mosquito net'
c.	o ~ u	*tsong*	/tsõː^{H}/ 'onion'	+	*zhing*	/sĩ^{H}qə^{H}/ 'field'	[tsṹːsĩ́ː]	'onion field'
d.	ɔ ~ ʊ	*tshogs*	/tsʰɔː^{H}/ 'assembly'	+	*gur*	/qʰuː^{LH}/ 'tent'	[tshʊ́ːqúː]	'assembly tent'
e.	ø ~ y	*bod*	/pʰøː^{LH}/ 'Tibet'	+	*lug*	/luː^{H}/ 'sheep'	[pʰỳːlúː]	'Tibetan sheep'

Table 3.4: Examples of regressive vowel height harmony in compound nouns (Dawson, 1980, p.69).

Second, Monich only cites a very limited number of vowel reduction examples as presented in Table 3.2. In fact, Denwood (1999) categorized these examples as stress-related alternations, but he was not convinced that there was stress or a metrical system in LT:

> At first sight, these [vowel alternations] look like stress-related alternations, with high prominence or stress on the first syllable and lower prominence at the second syllable. Such is indeed my impression from listening to the examples. However, the components of stress in the Lhasa dialect have never been satisfactorily analyzed, and the ac-

> counts of the various authors who have touched on the topic are not compatible enough to be reducible to a single statement [that LT has a productive stress/metrical system]. (Denwood, 1999, p.78)

An important point here is that vowel reduction is not pervasive in LT. The limited available data is not sufficient to establish that it occurs because the syllable is unstressed. The only other known cases of vowel reduction or laxing in unstressed syllables are given by Dawson (1980, p.23), who reported that two unstressed suffixes undergo vowel centralization: /-pa^{L}/ (narrative past verbal suffix), and /-la^{L}/ (locative suffix). Examples given by Dawson (1980, p.22) are reproduced in Table 3.1.

Stem		Suffix		Phonological output	Gloss
byas	/ɕʰɛː^{LH}/	*pa red*	/-pa^{L} + reː^{L}/	[ɕʰɛ̌ːpə̠rèː]	‘s/he did’
gnang	/naː^{H}/	*pa red*	/-pa^{L} + reː^{L}/	[náːpə̠rèː]	‘s/he did’ (H)
gzigs	/siː^{LH}/	*pa red*	/-pa^{L} + reː^{L}/	[sǐːpə̠rèː]	‘s/he looked, bought’ (H)
bod	/pʰøb^{LH}/	*la*	/-la^{L}/	[pʰø̌ːlə̠]	‘in, to Tibet’
rgya gar	/ka^{L}qaː^{H}/	*la*	/-la^{L}/	[kàqáːlə̠]	‘in, to India’
ri pin	/ri^{L}piː^{H}/	*la*	/-la^{L}/	[rìqíːlə̀]	‘in, to Japan’

Figure 3.1: Laxing of unstressed suffixes (Dawson, 1980, p.22).

Note, however, that even these two suffixes do not reduce in my recordings, or those provided by another author (Tournadre and Dorje, 2003). Furthermore, according to Dawson (1980, p.11, p.67), there are only two other environments where the alternation [a∼ə] can be observed: a) regressive raising (e.g., /naː^{H} + ku^{H}/ → [nə́ːkú] ‘s/he has yet done’), and (b) pre-bilabial laxing (e.g., /lap^{H}/→[lə́p] ‘study, teach’). This suggests that vowel laxing is most likely a fossilized alternation based on the stress pattern found in some LT words before tonogenesis (cf. Caplow, 2009).

There is thus no convincing phonological evidence to support the existence of stress or metrical structure in LT. In this dissertation, I argue that LT word tonology relies upon prosodic words to define the domain of tonal process, instead of invoking abstracts metrical feet or tonal feet. This approach is presented in the following section.

3.2 Word tonology of LT

LT word tonology derives surface tonal forms from underlying tonal representations. I claim that there are three basic lexical tones: L, LH, and H. H can be found in both stems and affixes, LH can only be found in stems, and L can only be found in affixes. Table 3.5 gives a contrastive minimal set of tones in LT[2].

Transcription	Segments	Underlying form	Gloss
spa	pá	/H/	'warrior'
sba	p/bǎ	/LH/	'cane'
pa	-pá	/H/	'nominalizer'
pa/ba	-p/bà-	/L/	'past tense verb linking particle' (cf. Denwood, 1999, p.142)

Table 3.5: Minimal pairs for tone contrast in LT

Evidence supporting the existence of two lexical tones in stems is well supported by the dominant two-tone analysis, as previously discussed in §2.1.1. I am in agreement with Duanmu (1990) and Yip (2002, p.196-199) that two lexical tones (LH, H) can be found in stems, and also more specifically, that the low register tone has an underlying form of /LH/. LH is unique, in that it can be treated either as a non-decomposable contour tone (LH), or be decomposed as two level tones (L and H), depending on the phonological context. The tonal status of the lexical tones found in affixes, however, is still in contention and will be discussed in §3.2.1. The word tonology of LT consists of two major components: a mechanism to define domains, and a system of rules operating on that domain. The first part is provided

[2]Based on the written form, I expect *spra* is pronounced as [drá]. That said, it is pronounced as [pá] by my LT speaker. According to the Tibetan legend, Tibetans are descendants of *spra* 'apes' living on the Tibetan plateau (Duff, 2015)

by prosodic structure, while the second part is what I called *Revised Lhasa Tone Rules* (RLS). I will describe and provide evidence for the LT prosodic structure in §3.2.2. Then, I will present the details of Revised Lhasa Tone Rules (§3.2.3), and describe their application in monosyllabic, disyllabic and polysyllabic words.

3.2.1 Tones of affixes

Previous analyses concerning the tones of affixes can be categorized into two main approaches. The first approach considers all affixes as atonal, such that they are either assigned a default L or have their pitch defined by the interpolation between the preceding and the following tone (Qu and Tan, 1983; Hu, 2002; Denwood, 1999). The second approach postulates that tonal suffixes bear an H while atonal suffixes are excluded from the tonal domain, and receive a L by default (Gong, 2014, p.6-7).

I propose that all affixes can either bear H or L. They have the same status as the lexical tones found in stems and can interact with adjacent lexical tones. The L found on affixes is neither an allotone of LH nor a default low because it interacts with other lexical tones in a phrasal construction (Lim, 2010a). For instance, the L genitive particle /-i^{L}/ in a genitive phrase construction merges with the preceding H (e.g., *phye* /p^{h}e^{H}/ 'powder'). After the merger, the segment of the genitive particle *ki* /-i/ is deleted, and compensatory lengthening takes place on the vocalic nucleus of the stem *phye* /p^{h}e/, (i.e., /p^{h}e + i/→[p^{h}eː]). Concurrently, the H of the stem merges with the L of the following genitive particle and forms a mid tone (/H + L/→[M]). Furthermore, the surface realization of the affixal L is distinctively different from LH, in that the L has a low level (or low-falling) pitch contour, as compared to the low-rising pitch contour of the LH. This coalescence process is shown in an elicited example (7).

(7) *phye ki phye*/$p^he^H + i^L + p^he^H$/ $\rightarrow$ [pēː pé] 'powder's powder'

3.2.2 Prosodic structure of LT

LT prosodic structure has five constituents: *a*) the mora (μ) *b*) the syllable (σ), *c*) the prosodic word (ω), *d*) the prosodic word group (Ω), and *e*) the intonational phrase (ι). Moras are organized into syllables, syllables are organized into prosodic words, prosodic words are organized into a prosodic word group, and finally prosodic word groups are organized into an intonational phrase, as shown in Figure 3.2. As there is no evidence for the prosodic constituency of the phonological utterance and the phonological phrase, the prosodic structure of LT does not adhere to the conventional Prosodic Hierarchy framework (cf. Nespor and Vogel, 1982; Selkirk, 1980; Peperkamp, 1999; Hildebrandt et al., 2010).

…ι… Intonational phrase
|
…Ω… Prosodic word group
|
…ω… Prosodic word
|
…σ… Syllable
|
…μ… Mora

Figure 3.2: The LT prosodic structure.

The main roles of these prosodic constituents are as follows: *a*) the mora is used to indicate the length or weight of the vowels, *b*) the syllable is the tone-bearing unit (TBU), *c*) the prosodic word delimits the domain of tonal process, *d*) the prosodic word group is the domain of downstep, and *e*) the intonational phrase is the domain of boundary tones.

A detailed discussion of the intonational phrase and its role in intonational phonology will be undertaken in Chapter 4. In the following discussion, I will focus on the mora, syllable, prosodic word, and prosodic word group, which operate at or below the word level.

3.2.2.1 The mora (μ)

The mora provides a way to represent the contrast of vowel length and syllable weight (Hyman, 1989; Hayes, 1989a; Zec, 2007; Yip, 2002; McCarthy and Prince, 1993). In LT, it is necessary to encode the vowel length contrast exemplified in Table 3.6. Note that in LT, the mora plays no role in phonological tonal implementation (duration has a moderate effect on the phonetic contour of tones, as we will see below).

	μ			$\mu\mu$		
a.	*spa*	pá	'warrior'	*spar*	pá:	'to light'
b.	*pha*	p^há	'father'	*phar*	p^há:	'that way'
c.	*ba*	p^hǎ	'cow'	*bar*	p^hǎ:	'in between'
d.	*sba*	bǎ	'cane'	*lba.ba*	bǎ:	'goitre'
e.	*ma*	má	'mother'	*mar*	má:	'butter'
f.	*rma*	mǎ	'sore'	*dmar*	mǎ:	'red/pink'

Table 3.6: Minimal pairs for vowel duration contrast.

3.2.2.2 The syllable (σ)

The syllable is widely accepted as the tone-bearing unit (TBU) in LT word tonology (Chang and Chang, 1964; Hari, 1977; Duanmu, 1992; Hu, 1980; Hu et al., 1982; Qu, 1981; Qu and Tan, 1983; Monich, 2011). There are alternate proposals which consider either the word (Sprigg, 1955) or the mora as the TBU (Yip, 2002; Lim, 2010a). Since rules defining the exact position of tones within words only need to refer to the syllable and since syllable weight does not seem to affect the complexity of tone contours in LT (e.g., *sba* /bǎ/ 'cane' vs. *lba.ba* /bǎː/ 'goitre') neither of these proposals are satisfactory and the syllable has to be retained as the TBU.

The alignment of the LH contour in disyllabic words can be used to illustrate that the syllable is the tone-bearing unit. An underlying LH monosyllabic word, such as *sba* /ba^{LH}/ 'cane,' has an [LH] phonological output with a rising pitch contour. However, an underlying /LH/ on the first syllable of a disyllabic prosodic word, such as *nyams myong* /(ɲam^{LH}.ɲuŋ^{LH})$_{\omega}$/ 'experience,' surfaces as a sequence of two level tones over two syllables [ɲam^{L}.ɲuŋ^{H}], after deletion of the non-initial tone (something I will come back to in §3.2.3).

3.2.2.3 The prosodic word (ω)

As in other languages, the prosodic word[3] is the most important constituent in word-level prosodic structure. As we will see in §3.2.3, it is the domain of application of the most important tone rules. Before addressing these rules, let us look at its formation. LT prosodic words are, at most, disyllabic (cf. Gong, 2014). Although LT is, historically, a monosyllabic language, the tendency for LT to move towards disyllabicity is evident, and a similar trend is observed among other Central Tibetan languages (Tournadre and Dorje, 2003, p.65). In the 90,000-word Tibetan database constructed by Kong (2012), there are 5,649 (6.2%) monosyllabic words and 43,000 disyllabic words (47.7%). LT now has a fair proportion of disyllabic morphemes, but most polysyllabic words are nonetheless compounds. What is crucial here is that there are relatively complex rules about how morphemes concatenate into prosodic words. These are morphological and semantic in nature:

Prosodic word boundaries align with morpheme boundaries: The LT prosodic word is sensitive to morphological structure. There is a strong isomorphism between

[3]Also known as the phonological word (Nespor and Vogel, 2007).

the prosodic word and the morphological structure of words, especially for lexical stems. For instance, *pus mo'i lha nga* ((pý.mǿ)$_{\omega}$(!há.ŋá)$_{\omega}$) 'kneecap,' a quadrisyllabic word, consists of two disyllabic morphemes: (pý.mǿ)$_{\omega}$ 'knee' and (háŋá)$_{\omega}$ 'basin.' Both of these words are disyllabic and can be used independently of each other. As these two disyllabic words are joined to form a quadrisyllabic word, the downstep of the second prosodic word indicates that they constitute two different prosodic words (cf. §3.2.2.4). The prosodic phrasing of this quadrisyllabic word completely coincides with its morphological structure.

An LT affix can either form its own monosyllabic prosodic word or forms a disyllabic prosodic word with an adjacent stem. For example, a trisyllabic word, *mang che ba* [mǎŋ.tɕé.wà], 'most,' consists of three morphemes: [mǎŋ] 'more,' [tɕé] 'very,' and [wà] '(nominalizer).' Each morpheme forms a prosodic word that yields the following prosodic phrasing: (mǎŋ)$_{\omega}$.(tɕé)$_{\omega}$.(wà)$_{\omega}$. It is interesting to note that the L of the suffix (i.e., /-wa^{L}/) is not deleted , which is evidence that it parsed as a prosodic word. This parsing is also observed with other L bearing affixes such as the nominalizer suffixes /-pa^{L}, -wa^{L}/, the negation prefix, /ma^{L}-/ and its variants /me^{L}- and mi^{L}-/. Other suffixes, however, do not form individual prosodic words. For example, H-bearing suffixes such as the word-final adjectivizer /-po^{H}/, appear to form a prosodic word with the preceding stem, e.g., *dkar po* [kár.pó] 'white'. As will be discussed in §3.2.3.2, the absence of downstep between the two syllables is evidence that there is no prosodic word boundary between the two syllables.

Prosodic word boundaries align with semantic grouping: Semantic grouping is another main motivation for prosodic phrasing. Semantic grouping refers to a phenomenon whereby elements that have a common core meaning are prosodically grouped. Semantic grouping overrides the non-isomorphism between prosodic phrasing and morphological construction (Ito and Mester, 2008).

In a polysyllabic word with reduplicated stems, the duplicated stems which are semantically identical and grouped together as a prosodic word. For instance, the prosodic phrasing for *nag thing thing* /naʔ^{LH}.t^{h}iŋ^{H}.t^{h}iŋ^{H}/ 'pitch black' is (nǎʔ)$_{\omega}$(t^{h}íŋt^{h}íŋ)$_{\omega}$. The word-initial morpheme [nǎʔ] is 'black,' and the reduplicated morpheme [t^{h}íŋ.t^{h}íŋ] is 'penetrating.' Here, the duplication indicates the intensity of the penetrating darkness.

Semantic grouping is not restricted to identical stems but also to stems that are in the same semantic category. For example, the polysyllabic word *pha ma gcig pa* [p^{h}á.má.tɕí.pá] 'same parent' has four morphemes: /p^{h}a^{H}/ 'father,' /ma^{LH}/ 'mother,' /tɕiʔ^{LH}/ 'one,' and /pa^{H}/ '(adjectivizer).' The prosodic phrasing for this word is (p^{h}á.má.)$_{\omega}$(tɕíʔ.pá)$_{\omega}$. I would like to focus on the first two morphemes, [p^{h}á] 'father' and [má] 'mother,' which form a prosodic word. Semantically, they are closely related kinship terms. Obviously, based on the compounding of these two stems, a new word meaning "parent" is synthesized. The meaning of the second prosodic word, (tɕíʔ.pá)$_{\omega}$, is 'same'. As both of these compound words are used as independent disyllabic words, the prosodic phrasing maintains the structure of the compounds to preserve their meanings.

Semantic grouping can also be based on a culture-specific association. For example, *skye rga na 'chi* [k^{j}é.gá.nà.tɕ^{h}í] 'life' or 'cycle of life,' a common Tibetan phrase, consists of four morphemes: /k^{j}e^{H}/ 'birth,' /ga^{LH}/ 'old,' /na^{LH}/ 'sickness,' and /tɕ^{h}i^{H}/ 'death' [4] Nonetheless, it has the same prosodic phrasing (i.e., two disyllabic prosodic words) as the previous example: (k^{j}é.gá)$_{\omega}$(nà.tɕhí)$_{\omega}$. In this case, these prosodic phrasings are defined by semantic groupings according to the Buddhist tradition of cause and effect. In short, semantic grouping can also have an isomorphic relationship with prosodic phrasing, which overrides the isomor-

[4] This is a phrase that originated from the first of the Four Noble Truths of Buddha teaching (Lama, 1997).

phic relation of the morphological structure. Table D.1 and Table D.2 in Appendix E provide a summarized morphological analysis and prosodic phrasing of the LT polysyllabic words encountered in this dissertation.

3.2.2.4 The prosodic word group (Ω)

There is no empirical evidence that there is a prosodic domain in LT that matches syntactic XPs, like the phonological phrase proposed in standard models (cf. Selkirk, 1995; Nespor and Vogel, 2007). Furthermore, the absence of phonological phrases in Sino-Tibetan and South-East Asian languages are not unusual (cf. Hildebrandt et al., 2010). Thus, the prosodic word group (Ω) is the intermediate prosodic level between the intonational phrase and prosodic words in LT (Vigário, 2010). It is a compound-like structure that constitutes the domain for the implementation of downstep. This is, thus, a deviation from the Prosodic Hierarchy framework. Similarly, Hall and Hildebrandt (2008) have also identified grammatical words (*gword*) that consist of two prosodic words in Kyirong Tibetan.

Two polysyllabics words are chosen to illustrate the formation of prosodic word groups: *skye rga na 'chi* [k^jé.gá.nà.$tɕ^h$í] 'cycle of life' and *ka ra dkar po* [ká.rá.kár.pó] 'white sugar'. The former consists of four morphemes (*skye* /k^je^H/ 'birth,' *rga* /ga^{LH}/ 'old,' *na* /na^{LH}/ 'sickness,' *'chi* /$tɕ^hi^H$/ 'death') and the latter three (*ka ra* /$kara^H$/ 'sugar', *dkar* /kar^H/ 'white', and *po* /-po^H/ 'adjectivizer'). Both of them form prosodic word groups with two disyllabic prosodic words: $((k^j$é.gá.$)_\omega($nà.$tɕ^h$í$)_\omega)_\Omega$ and $(($ká.rá$)_\omega($!kár.pó$)_\omega)_\Omega$. The prosodic word group (Ω) boundaries naturally align with the morphological boundaries of the polysyllabic word.

3.2.3 Revised Lhasa Tone Rules (RLS)

In this section, I will present the *Revised Lhasa Tone Rules* (RLS) that govern tone assignment in LT prosodic word and show an example of the derivation of monosyllabic, disyllabic, and polysyllabic words. These revised tone rules are based on the Duanmu (1992) Lhasa Tone Rules (LS), which are an autosegmental analysis. I have extensively discussed the *Lhasa Tone Rules* in §2.1.3. The RLS have three advantages over Duanmu's original *Lhasa Tone Rules* (see §2.1.3.1): *a*) The prosodic word (which has up to two syllables) delimits the domain of tonal processes, *b*) allows up to two tones per syllable (i.e. /LH/ can associate with one syllable), and *c*) predicts phonological outputs on polysyllabic words more effectively.

I propose the following rewording of Duanmu's *Lhasa Tone Rules* (LS):

(8) Revised Lhasa Tone Rules (RLS)

a. *Prosodic word formation* (W): Form prosodic words of a maximum length of two syllables, based on morphemic boundaries and semantic symmetry.

b. *Non-initial tone deletion* (TD): Delete all non-initial tones (where LH is treated as a single unit) within a prosodic word (ω).

c. *Tone association* (TA): Associate tones to syllables one-to-one and left-to-right within a prosodic word wherever there is an available TBU (where LH is decomposable). LH, however, can associate with a monosyllabic prosodic word (where LH is treated as a single unit).

I will proceed to demonstrate the derivation of the phonological output (PO) of monosyllabic words, disyllabic words, and polysyllabic words from underlying representations (UR) based on the application of RLS in the following sections.

3.2.3.1 RLS in monosyllabic and disyllabic words

In LT, each monosyllabic word forms a monosyllabic prosodic word (application of RLS prosodic word formation (W)). For example, *pha* /p^ha^H/ 'father' and *ba*

/p^{h}a^{LH}/ 'cow' form the prosodic words (p^{h}á)$_{\omega}$ and (p^{h}ǎ)$_{\omega}$. In a monosyllabic word, the rule of non-initial tone deletion (R0) does not apply; the rule of tone association (R1) links H and LH with the syllable. The phonological outputs (PO), thus, can either be H or LH. I adopt Monich's (2011) proposal that tonal crowding is allowed when the LH is associated with a single TBU, in contrast with Duanmu's Lhasa Tone Rules that forbid tonal crowding. The reason for this decision is that, as I will show in §3.3, the LH tone is systematically realized as a rising pitch contour in monosyllables, regardless of the duration of their rhymes. Derivations of monosyllabic words are shown in Figure 3.3.

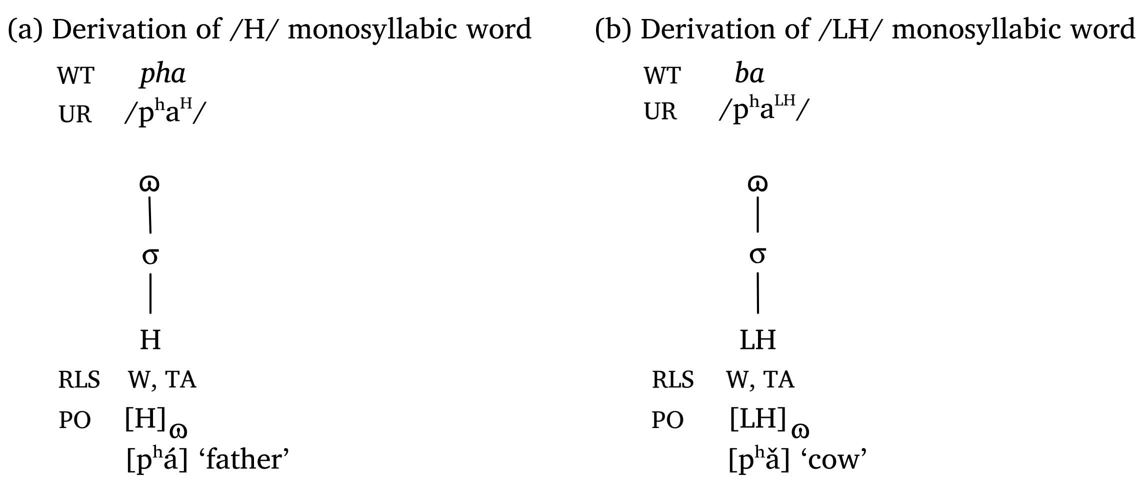


Figure 3.3: Derivation of monosyllabic words.

I have presented the list of disyllabic words of Qu and Tan (1983, p.35) in Table 3.10 to illustrate the application of RLS. Four possible underlying representations: (/H.H/, /H.LH/, /LH.H/, /LH.H/) only yield two phonological outputs: (H.H, L.H) regardless of rhyme structure.

No.	Rhyme length	Tone Sandhi			WT	Segments	Gloss
Group A							
1.	s.s	/H.H/	→	[H.H]	*shug pa*	[(ɕúʔ.pá)$_{\omega}$]	‘cypress’
2.	s.s	/H.LH/	→	[H.H]	*mna’ ma*	[(ná.má)$_{\omega}$]	‘bride’
3.	s.l	/H.H/	→	[H.H]	*chu shel*	[(tɕʰú.ɕél)$_{\omega}$]	‘crystal’
4.	s.l	/H.LH/	→	[H.H]	*ha yang*	[(há.jáŋ)$_{\omega}$]	‘aluminium’
Group B							
9.	l.s	/H.H/	→	[H.H]	*khang pa*	[(kʰáŋ.pá)$_{\omega}$]	‘house’
10.	l.s	/H.LH/	→	[H.H]	*sla nga*	[(láŋ.ŋá)$_{\omega}$]	‘frying pan’
11.	l.l	/H.H/	→	[H.H]	*bsam tshul*	[(sám.tɕár)$_{\omega}$]	‘opinion’
12.	l.l	/H.LH/	→	[H.H]	*shing sdong*	[(ɕíŋ.tóŋ)$_{\omega}$]	‘tree’
Group C							
5.	s.s	/LH.H/	→	[L.H]	*bdag po*	[(tàʔ.pó)$_{\omega}$]	‘master’
6.	s.s	/LH.LH/	→	[L.H]	*tho rdo*	[(tò.ró)$_{\omega}$]	‘pile of rocks’
7.	s.l	/LH.H/	→	[L.H]	*me phor*	[(mè.pór)$_{\omega}$]	‘fire pan’
8.	s.l	/LH.LH/	→	[L.H]	*ja ril*	[(tɕà.ríl)$_{\omega}$]	‘bowl shaped tea brick’
Group D							
13.	l.s	/LH.H/	→	[L.H]	*zam pa*	[(sàm.pá)$_{\omega}$]	‘bridge’
14.	l.s	/LH.LH/	→	[L.H]	*gdung ma*	[(tùŋ.má)$_{\omega}$]	‘roof beam’
15.	l.s	/LH.H/	→	[L.H]	*ngal rtsol*	[(ŋàl.tsǿ)$_{\omega}$]	‘labour’
16.	l.l	/LH.LH/	→	[L.H]	*nyams myong*	[(ɲàm.ɲúm)$_{\omega}$]	‘experience’

Table 3.7: Disyllabic pitch pattern of LT adapted from Qu and Tan (1983) and Duanmu (1992). The first syllable of Group A and C has an open syllable or glottal coda (s), but Group B and D has a closed syllable (l).

All of the disyllabic grammatical words in this paradigm form disyllabic prosodic words (Rule of Prosodic Word Formation). They illustrate the rule of deletion of non-initial tone, in which the lexical tones (either H or LH) of non-initial syllables are deleted (cf. Duanmu, 1992). I have chosen a word with initial LH, *gdung ma* /tuŋ^{LH}.ma^{LH}/ ‘roof beam’, as an example. The two tones of the stems of this word both have an underlying representation (UR) of /LH/, the second LH is deleted by the non-initial tone deletion rule (R0), as shown in Figure 3.4. The tone spreading rule (R1) then associates the L with the first syllable and the H with the second, yielding [tuŋ^{L}.ma^{H}].

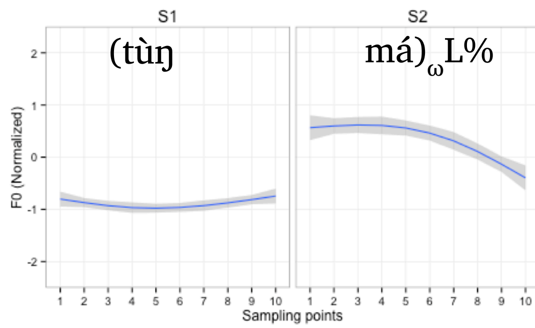


Figure 3.4: Pitch contour and prosodic word boundary of *gdung ma* /tuŋ^{LH} + ma^{LH}/→ [tùŋ.má] ‘roof beam’.

When a disyllabic word starts with a H lexical tone, the H is realized on both syllables, as can be seen in examples in Groups A and B in Table 3.10. I have chosen *mna’ ma* /na^{H}.ma^{LH}/ ‘bride’ to illustrate this. The word has an underlying representation of /H.LH/ and a phonological output of [H.H]. The pitch trace in Figure 3.5 shows that the /LH/ is not realized. This suggests that the tone of the second syllable is deleted and the word-initial H spreads to the second syllable.

In Figure 3.6, I have provided a full derivation of the phonological output of *mna’ ma* [ná.má] ‘bride’ and *gdung ma* [tùŋ.má] ‘roof beam’ based on RLS. These two phonological outputs, [H.H] and [L.H.], represent the two main types of LT disyllabic words. I have omitted the interaction between lexical tone and intonational tone (i.e. phrase-final falling induced by the L%), which will be further discussed in Chapter 4 (evidence for it can be found in Figures 3.4 and 3.5).

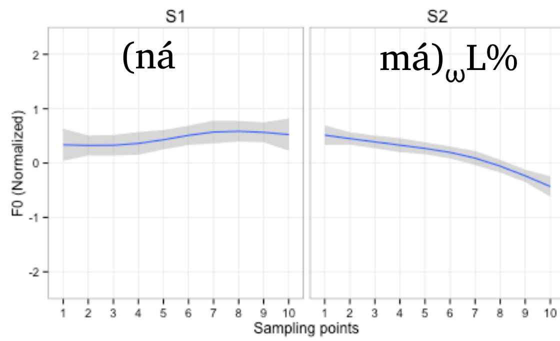


Figure 3.5: Pitch contour and prosodic word boundary of *mna' ma* /na^{H} + ma^{LH}/→[ná.má] 'bride'.

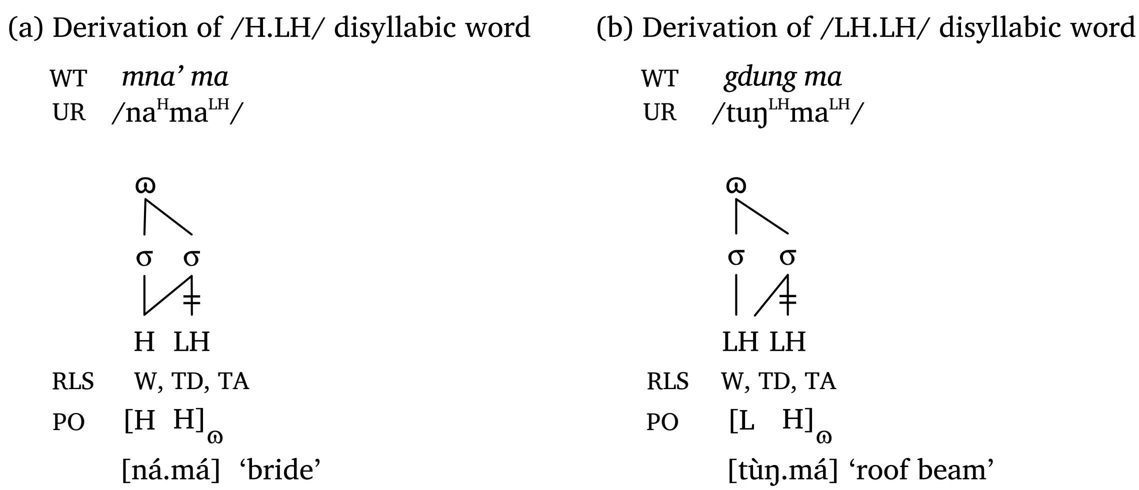


Figure 3.6: The derivation of disyllabic words based on RLS.

3.2.3.2 Application of RLS in polysyllabic words

I propose that the RLS have to be enriched to include a downstep rule that applies inside the prosodic word group. When two high tones are adjacent in a prosodic word group, the second one must be downstepped to a !H, a process that can be attributed to the Obligatory Contour Principle (Yip, 1988; Odden, 1988; McCarthy, 1986). This process is illustrated in Figure 3.7 and can be formulated as in (9).

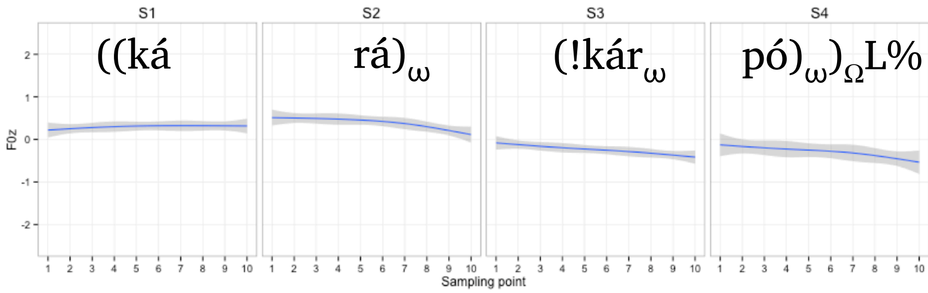


Figure 3.7: An example of downstep in a prosodic word group.

(9) *Downstep* (D): If two H tones are adjacent to each other across a prosodic word boundary within a prosodic word group, the second H is downstepped.

In what follows, I will present the derivation of representative examples of trisyllabic words and quadrisyllabic words. Although I am mostly extending principles proposed in previous studies, there has never been any attempt, to my knowledge, to account for the tones patterns of such words.

Examples (a) and (b) in Figure 3.8, *bka' drin can* /ka^{H}.tin^{H}.tɕeː ^{H}/ 'the kind one' and

nag thing thing /naʔ^{H}.tʰiŋ^{H}.tʰiŋ^{H}/ 'pitch black,' have the same underlying tonal representation but have different phonological outputs because the downstep occurs on the third and second syllables, respectively. I attribute these different phonological outputs to different prosodic word parsings, as shown in Figure 3.8. Examples (c) and (d), *snying rje po* [níŋ.tɕè.pó] 'beautiful' and *bka' gnang ba* [ká.nàŋ.wá] 'to order' in Figure 3.8, show that the initial H does not spread to the following syllable, and that the word-medial LH is preserved, indicating that the initial syllable of these words form their own prosodic word. In *snying rje po* [níŋ.tɕè.pó] 'beautiful', the word-medial LH is decomposed into L and H, where the latter spreads to the third syllable after the tone of the third syllable is delinked. This analysis is sup-

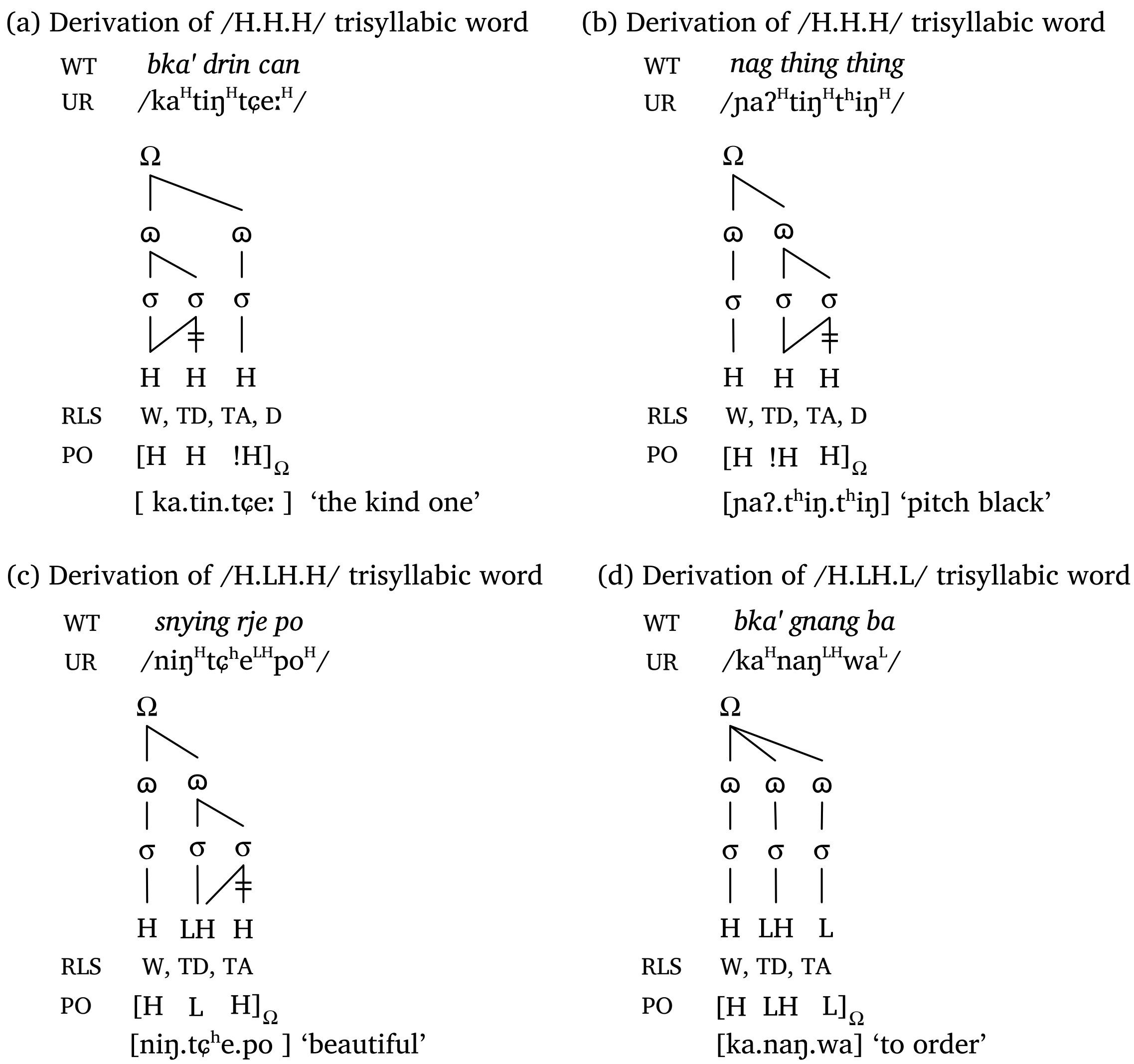


Figure 3.8: LT trisyllabic words with word-initial H.

ported by the surface realization of two plateaux in this disyllabic prosodic word. In contrast, the word-medial LH of *bka' gnang ba* [ká.nǎŋ.wá] 'to order' does not decompose because the second syllable is parsed as its own monosyllabic prosodic word.

The data in Figure 3.8 does not confirm the maximum disyllabicity and push-out effect proposed by Gong (2014), and discussed in §2.1. Gong proposes that the tonal domain of a trisyllabic word is restricted to the initial two syllables, and the third syllable receives an automatic low tone. This would lead us to expect *bka' drin can* /ka^{H}.tin^{H}.tɕeː^{H}/ 'the kind one' to surface as [ká.tín.tɕèː], which is an incorrect prediction. As presented in the derivation above, the RLS predicts the realization of the final tone as a downstepped H. Gong's proposal also fails to predict the derivation and surface realization of quadrisyllabic words, discussed below.

Figure 3.9 presents examples of phonological derivations in trisyllabic words with phrase-initial LH lexical tone. These examples are used to illustrate the tonal interaction of non-decomposed and decomposed LH tone with adjacent H or LH. I argue that the non-decomposed LH of a monosyllabic prosodic word, as in example (a) *las sla po* [lěː.lá.pó] 'easy' and (b) *ngo shes pa* [ŋǒ.ʃéː.pà] 'acquaintance' in Figure 3.9, is not able to downstep the following H. The LH of these monosyllabic prosodic words is evidenced by the presence of a continuous rising pitch contour in the output (see Appendix C C.4). However, example (d), *gzhes ka rtse* [ʃì.gá.!tsé] 'Shigatse' shows that the decomposed H of LH can downstep the H of the third syllable. This suggests that LH is ambiguous when it comes to decompositionality. It is a unitary tone that does not trigger downstep when it is on a single syllable. However, after spreading over two syllables, its two parts become independent and H can trigger downstep. To my knowledge, no other examples of positionally conditioned decompositionality have ever been reported in the literature.

Figure 3.10 shows examples of quadrisyllabic words accompanied by their phonological derivations. Examples (a) and (b), *pus mo'i lha nga* [pý.mǿ.há.ŋá] 'knee cap' and *skal ba bzang po* [kèl.bá.ʃàn.pó] 'excellent fortune', are parsed into two disyllabic prosodic words. In example (a), *pus mo'i lha nga* [pý.mǿ.!há.ŋá] 'knee cap', the prosodic word boundary can be be diagnosed because of the down-

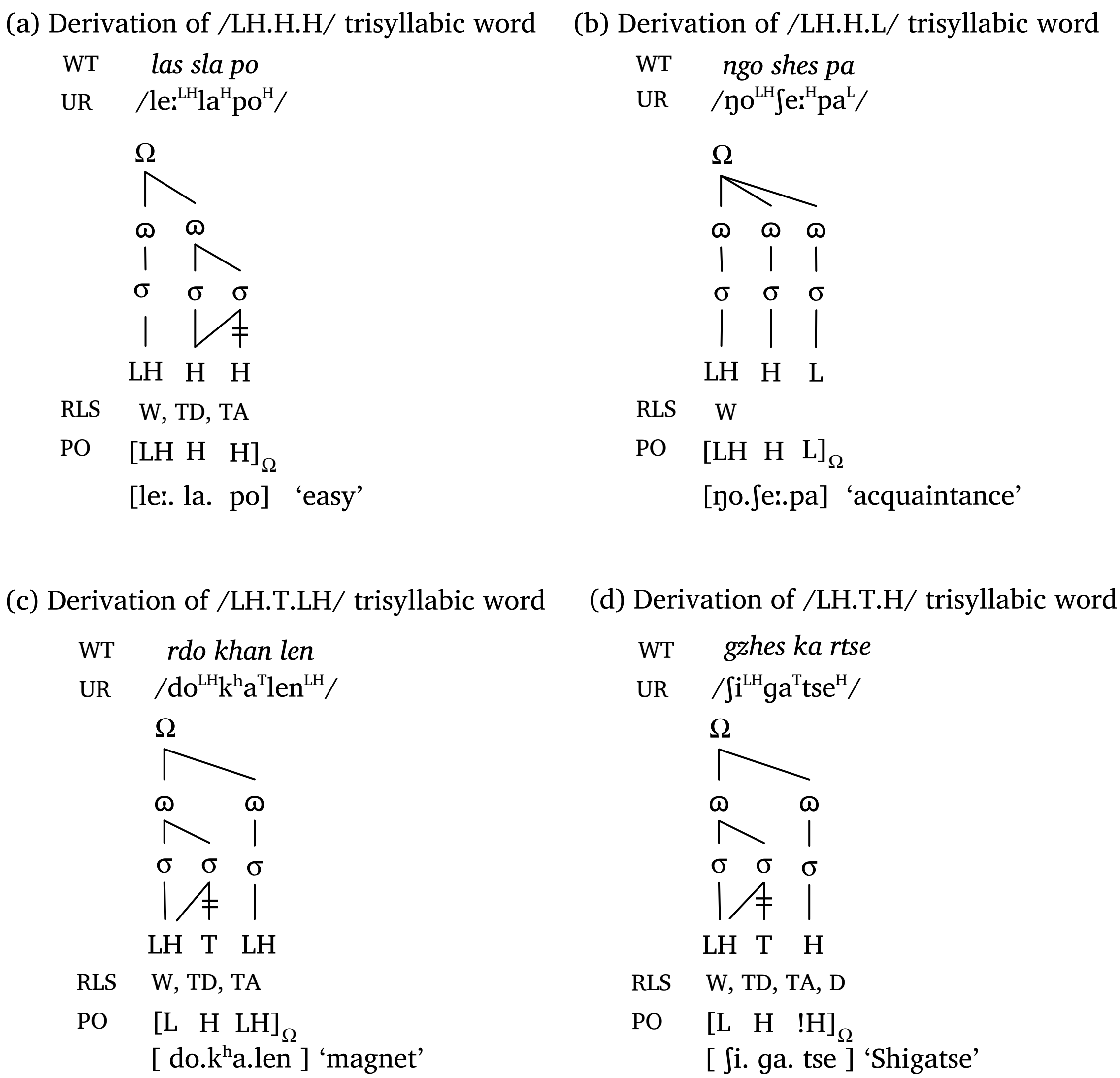


Figure 3.9: LT trisyllabic word examples with word-initial LH. (Note: The unknown underlying form of a non-initial tone is represented by “T”.)

step. In example (b), *skal ba bzang po* [kèl.bá.ʃàn.pó] ‘excellent fortune’, the preservation of the LH tone of the second prosodic word shows that the rule of non-initial tone deletion (R0) does not apply inside the prosodic word group. Likewise, in example (c), the second prosodic word-initial LH of *tig.pa’ra.tsa* [dì.pá.rà.tsá] ‘scorpion’ is preserved, leading to a [L.H.L.H] pitch pattern. Example (d), *dpag tu med* *pa* [páʔ.dǔ.mè.pà] ‘unfathomable’ is an example of a prosodic phrasing in which each morpheme within the quadrisyllabic word forms a prosodic word.

The tonal patterns of LT words can therefore be captures by an interaction of prosodic parsing and relatively simple tone rules. In the next section, I will look at the phonetic realization of the surface tonal representations obtained from the RLS.

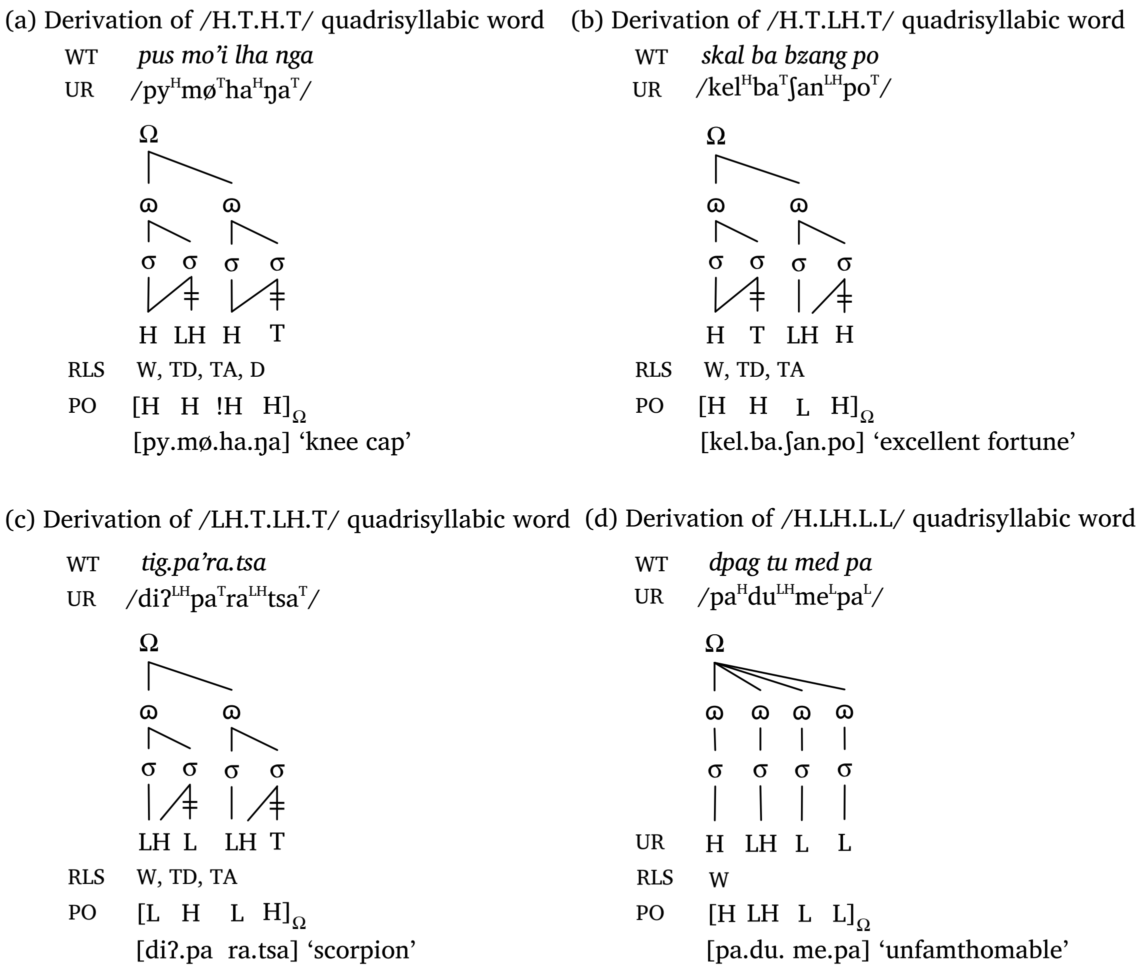


Figure 3.10: LT quadrisyllabic word examples. (Note: The unknown underlying form of a non-initial tone is represented by “T”.)

3.3 Experiment I: Surface realization of the tones of LT words

The previous section outlines the derivation of the phonological outputs of LT monosyllabic, disyllabic, and polysyllabic words based on underlying representations containing the tones /L, LH, H/. This section describes an experiment that investigates the surface realization of underlying forms with these tones, and the interaction of surface tones with different syllable type. The research questions for this experiment are:

(10) a. What is the surface realization of tone in LT words and does it match the description laid out in previous sections?
b. Is there any duration difference between different types of syllables and does it affect tonal patterns?

3.3.1 Methods

3.3.1.1 Subjects

The subjects who participated in this experiment are from the Tibetan communities in Ottawa and Toronto. There are two subjects from Ottawa and six from Toronto: seven male speakers and one female speaker. I originally contacted seven female LT speakers, but only one female LT speaker was willing to participate in the interview [5]. The participants were assigned a prefix corresponding to their gender ("M" for male, "F" for female), followed by a sequential numeral designation, such as "M1" or "F1." In the following table, I summarize the linguistic background of all

[5] My initial impression of these elderly female subjects is that they were very humble and traditional. These characteristics were clearly manifested through their demeanour around men, and they also tended to be reserved and refrained from actively participating in casual conversations during our visit. Many of them politely and tactfully declined my interview in front of their husbands, who participated in this experiment. I suspect some of them considered my linguistics interview to be official business since I presented myself as a researcher from the University of Ottawa; the female subjects might have thought that as this interview was official business, it should be handled by their husbands. Moreover, my language consultant is a well-known community leader serving the Tibetan community in Canada. In short, I need to rethink my strategy for future research with female speakers so that they will feel more comfortable to participate. Hopefully, this social dynamic can be further investigated in future sociolinguistics studies.

participants in the study.

Subjects	Age	1st L1	2nd languages (L2)
M1	60s	LT	English, Japanese, Hindi, Nepali,
M2	60s	LT	English, Hindi
M3	40s	Ütsang	English, Hindi, Nepali, Punjabi
M4	60s	LT	English, Danish
M5	70s	LT	English, Hindi, Nepali
M6	70s	LT	English, Hindi, Nepali, Punjabi
M7	50s	LT	English, Hindi, Nepali, Mandarin
F1	70s	LT	Hindi, Nepali, Punjabi

Table 3.8: Linguistics background of LT participants.

All participants are multilingual[6] but identified LT as their first language (L1); they originated mainly from the Central Tibetan area and had lived in or around Lhasa City. At least one of their parents was also an LT or *Ütsang* speaker. LT is a dialect of Central Tibetan *Ütsang*; thus, it is intelligible to *Ütsang* speakers as well. Subject M3, however, identified himself as a *Ütsang* Central Tibetan rather than Lhasa Tibetan (LT) speaker. He shows some minor dialectal differences with other subjects that will be identified during the discussion of results. All subjects had received some form of instruction in Classical Tibetan from a monastery-based educational institution. Most of them are highly proficient in literary Tibetan, although LT is not necessarily the medium of instruction in traditional monastic schools. The local Tibetan dialect that is used for instruction depends on the location

[6]These subjects have a diverse linguistic background; that is, they are proficient in several other languages (L2). Some of the languages were acquired through contact and not necessarily through formal instruction. These subjects are part of the Tibetan diaspora from India and Nepal, who later immigrated to Canada. When they were in India or Nepal, some of them had the opportunity to receive English education at the primary or secondary school level. One of the subjects even attended high school and received a university degree in the United Kingdom. All subjects, thus, can communicate in English except F1. Most of them had lived and worked in Northern India and Nepal, where they acquired proficiency in Hindi, Punjabi, and Nepali. There are two subjects who are also fluent in other languages, Japanese and Danish, respectively, because they received their tertiary education in Japan and Denmark. Participant M7 can also speak some rudimentary Mandarin.

of the monastic school. My main language consultant considered all subjects as native LT speakers.

3.3.1.2 Stimuli

For this experiment, I used three groups of stimuli: monosyllabic words, disyllabic words, and polysyllabic words.

Monosyllabic words with four different onsets, /p, p^h, b, m/[7], and six rhymes, /a, aː, aʔ, aŋ, ar, al/[8], are included in the present experiment. The six sets of monosyllabic stimuli are shown in Table 3.9. Many more rhyme types can be found in LT (see Qu and Tan, 1983), but the ones included here are those that are relatively common. A word with a /p/ onset may only bear an H lexical tone (e.g., [pá]), while one with a [b] onset may only bear an LH lexical tone (e.g., [bǎ]). A word with a /p^h/ or /m/ onset can carry either an H (e.g., [p^há]) or LH (e.g., [p^hǎ]) lexical tone.

For disyllabic words, I used the LT disyllabic word paradigm of Qu and Tan (1983) previously discussed in §3.2.3.1 as stimuli. I have reproduced it in abridged form in Table 3.10.

[7]In the current analysis, I consider that the voicing of onset stop consonants is phonologically redundant, as there appears to be free variation among the speakers. A learned LT speaker will tend to voice onset stop consonants in agreement with the orthography. In casual speech, LT is losing the voicing contrast of onset stop consonants, as reported in previous research (Hu, 2002).

[8]Monosyllabic words with *ak* rhyme in written form can surface as [aʔ] or [ak̚]. I have used [aʔ] in current transcription. Furthermore, long vowel /aː/ are produced by speakers that contrast between the written ([ar]) and colloquial pronunciation ([aː]). Some speakers are not aware of this contrast and I do not include these data in my analyses.

	I			**II**		
a.	*spa*	[pá]	‘warrior’	*spar*	[páː]	‘to light’
b.	*pha*	[pʰá]	‘father’	*phar*	[pʰáː]	‘that way’
c.	*bha*	[pʰǎ]	‘cow’	*bar*	[pʰǎː]	‘in between’
d.	*sba*	[bǎ]	‘cane’	*lba.ba*	[bǎː]	‘goitre’
	III			**IV**		
a.	*spang*	[páŋ]	‘lap’	*bag*	[páʔ]	‘barley dough’
b.	*phang*	[pʰáŋ]	‘spindle’	*phag*	[pʰáʔ]	‘pig’
c.	*bhang*	[pʰǎŋ]	‘foot race’	*bhag*	[pʰǎʔ]	‘feeling’
d.	*bang*	[bǎŋ]	‘beer waste’	*sbag*	[bǎʔ]	‘Tibetan mahjong’
	V			**VI**		
a.	*spar*	[pár]	‘to light’ (lit.)	*dpal*	[pál]	‘prominent’
b.	*phar*	[pʰár]	‘that way’ (lit.)	*phal*	[pʰál]	‘common’
c.	*bar*	[pʰǎr]	‘in between’ (lit.)	*bal*	[pʰǎl]	‘wool’
d.	*lba.ba*	[bǎr]	‘goitre’ (lit.)	*bal*	[bǎl]	‘plenty’
	VII			**VIII**		
a.	*ma*	[má]	‘mother’	*rma*	[mǎ]	‘sore’
b.	*mar*	[máː]	‘butter’	*dmar*	[mǎː]	‘red’
c.	*mang*	[máŋ]	‘many’	*dmangs*	[mǎŋ]	‘commoner’
d.	*mag*	[máʔ]	‘groom’	*dmag*	[mǎʔ]	‘war’
e.	*mal*	[mál]	‘one’s own place’	*smal*	[mǎl]	‘name of a star’
f.	*mar*	[már]	‘butter’ (lit.)	*dmar*	[mǎr]	‘red’ (lit.)

Table 3.9: Monosyllabic words (bilabial onsets and various codas).

Finally, for words of three or more syllables, I selected a number of stimuli for this experiment to try to give a balanced sample of the combinations of lexical tone found in polysyllabic words. The pitch pattern of these combinations will allow us to examine the interaction between various lexical tones and prosodic boundaries. These polysyllabic stimuli are adopted from wordlists prepared by Qu and Tan (1983) and from the *The Iluminator Tibetan-English Encyclopedic Dictionary* (Duff, 2015). I have treated each polysyllabic word as a single prosodic word group (Ω) and defined prosodic words (ω) internal to the prosodic word groups according

No.	Rhyme length	WT	Segments	Gloss
Group A				
1.	s.s	*shug pa*	[(ɕúʔ.pá)$_{\omega}$]	‘cypress’
2.	s.s	*mna’ ma*	[(ná.má)$_{\omega}$]	‘bride’
3.	s.l	*chu shel*	[(tɕ^{h}ú.ɕél)$_{\omega}$]	‘crystal’
4.	s.l	*ha yang*	[(há.jáŋ)$_{\omega}$]	‘aluminium’
Group B				
5.	l.s	*khang pa*	[(k^{h}áŋ.pá)$_{\omega}$]	‘house’
6.	l.s	*sla nga*	[(láŋ.ŋá)$_{\omega}$]	‘frying pan’
7.	l.l	*bsam tshul*	[(sám.tɕár)$_{\omega}$]	‘opinion’
8.	l.l	*shing sdong*	[(ɕíŋ.tóŋ)$_{\omega}$]	‘tree’
Group C				
9.	s.s	*bdag po*	[(tàʔ.pó)$_{\omega}$]	‘master’
10.	s.s	*tho rdo*	[(tò.ró)$_{\omega}$]	‘pile of rocks’
11.	s.l	*me phor*	[(mè.pór)$_{\omega}$]	‘fire pan’
12.	s.l	*ja ril*	[(tɕà.ríl)$_{\omega}$]	‘bowl shaped tea brick’
Group D				
13.	l.s	*zam pa*	[(sàm.pá)$_{\omega}$]	‘bridge’
14.	l.s	*gdung ma*	[(tùŋ.má)$_{\omega}$]	‘roof beam’
15.	l.s	*ngal rtsol*	[(ŋàl.tsǿ)$_{\omega}$]	‘labour’
16.	l.l	*nyams myong*	[(ɲàm.ɲúm)$_{\omega}$]	‘experience’

Table 3.10: Disyllabic pitch pattern of LT adapted from Qu and Tan (1983) and Duanmu (1992); the first syllable of Group A and C has an open syllable or glottal coda (l), but Group B and D have a closed syllable (s).

to the known morphology and semantic grouping of LT words (cf. §3.2.2.3). The morphemes of these polysyllabic words are presented in Appendix D. The trisyllabic and quadrisyllabic stimuli are presented in Tables 3.11 and 3.12, grouped by prosodic phrasing.

WT	Segments	Gloss
$[((\sigma\sigma)_\omega(\sigma)_\omega)_\Omega]$		
rdo khab len	[dò.k^{h}áp.lěn]	‘magnet’
phjig ling ba	[tɕí.líŋ.wà]	‘foreigner’
bka’ drin can	[ká.tín.tɕéː]	‘the kind one’
katora	[ká.tó.rǎ]	‘red copper basin’
ka tsha po	[ká.ts^{h}á.pó]	‘urgent’
skad cig ma	[kéʔ.tɕíʔ.mǎ]	‘a moment’
potala	[pó.tá.là]	‘Potala palace’
gzhes ka rtse	[ʃì.ká.tsé]	‘Shigatse’
phyem ma leb	[tɕé.má.lěb]	‘moth’
$[((\sigma)_\omega(\sigma\sigma)_\omega)_\Omega]$		
skad chen po	[kéʔ.tɕ^{h}én.pó]	‘loud, or noisy’
skad gnyis pa	[kéː.ɲí.pà]	‘bilingual’
les sla po	[lěː.lá.pó]	‘easy’
nak thing thing	[nǎʔ.t^{h}íŋ.t^{h}íŋ]	‘pitch black’
ning che po	[níŋ.tɕ^{h}é.pó]	‘beautiful’
sngo thing thing	[ŋó.t^{h}íŋ.t^{h}íŋ]	‘bluish’
ser hang hang	[sér.há.há]	‘yellowish’
$[((\sigma)_\omega(\sigma)_\omega(\sigma)_\omega)_\Omega]$		
rkang pa’i nyva	[káŋ.péː.ɲǎ]	‘calf’
rkyal rgyag pa	[k^{j}él.g^{j}ǎʔ.pà]	‘to swim’
ma ’ongs pa	[mà.ǒŋ.pà]	‘future’
mi dge ba	[mì.gé.wà]	‘non-virtue’
mi bzod pa	[mì.ʃǿː.pà]	‘unbearable’
mi rtag pa	[mì.ták.pà]	‘impermanence’
mang che ba	[mǎŋ.tɕ^{h}é.wà]	‘most’
ngo shes pa	[ŋǒ.ʃé.pà]	‘acquaintance’
shel kara	[ʃél.kárá]	‘rock sugar’

Table 3.11: LT trisyllabic stimuli.

WT	Segments	Gloss
$[((\sigma\sigma)_{\omega}(\sigma\sigma)_{\omega})_{\Omega}]$		
tig pa'ra tsa	[dìʔ.pá.rà.tsá]	'scorpian'
dkrug dkrug gtong ba	[tɕúʔ.tɕúʔ.tóŋ.wà]	'stirring'
dka' las khag po	[ká.lék.kʰáʔ.pó]	'difficult'
dkar ra dkar po	[ká.rá.kár.pó]	'white sugar'
rkang brgya lag brgya	[káŋ.kʲá.làk.kʲá]	'centipede'
dkar po dkar rkyang	[kár.pó.kár.kʲáŋ]	'pure white'
skye ba snga ma	[ké.wá.ŋá.má]	'former life'
skyes bu dam pa	[kʰí.bú.dàm][pá]	'saint'
skye rga na 'chi	[kʲé.gá.nà.tɕhí]	'life/cycle of life'
pha ma gcig pa	[pʰá.má.tɕíʔ.pá]	'same parent'
pu pu khu shud	[pú.pú.kʰú.ʃýʔ]	'hoopoe'
pus mo'i lha nga	[pý.mǿ.há.ŋá]	'knee cap'
cema kara	[tɕè.má.kárá]	'coarse sugar'
cem cel ha mo	[tɕém.tɕél.hámó]	'butterflies'
$[((\sigma\sigma)_{\omega}(\sigma)_{\omega}(\sigma)_{\omega})_{\Omega}]$		
'bu cha ga pa	[bù.tɕʰá.gǎ.pà]	'grasshopper'
dbjar rtsha dgun 'bu	[dʑár.tsá.gùn.bú]	'herbal medicine'
skad cha bshad pa	[ké.tɕʰá.ʃé.pà]	'to talk'
skal ba bzang po	[kél.bá.ʃàŋ.pó]	'excellent fortune'
skyid po gtong ba	[kí.pú.tóŋ.wà]	'to be happy'
kyi hud 'don pa	[kʲí.hýʔ.dǒn.pà]	'Oh! oh! wailing'
ku co sgrog pa	[kú.tɕó.drǒ.pà]	'to make noise'
$[((\sigma)_{\omega}(\sigma)_{\omega}(\sigma)_{\omega}(\sigma)_{\omega})_{\Omega}]$		
dpag tu medpa	[páʔ.dǔ.mè.pá]	'unfathomable'

Table 3.12: LT quadrisyllabic stimuli.

3.3.1.3 Procedure

All recordings were made in a quiet place at the participants' homes. Participants were given the wordlist and were instructed to read the words out loud, using a regular intonation at a normal rate. Each word was repeated three consecutive times with pause in between repetition, resulting in three tokens (T1, T2, T3) for each word. I intentionally did not use a frame sentence to detect and capture the intonational pattern that the participant might use during elicitation. All these tokens were recorded in a mono WAV format with a sampling rate of 44.1 kHz and a bit depth of 16.

3.3.1.4 Equipment

All words were recorded using a Shure Beta53 omnidirectional measurement microphone with an Apogee Duet external mic preamplifier and analog-to-digital converter (ADC). The preamplifier was connected to a MacBook Pro via a USB cable. The words were recorded in Praat (Weenink and David, 2017) after selecting the Apogee Duet as the audio input. I also monitored the recording session through headphones. The microphone gain was preset before testing. I did not change the gain during the recording unless the speaker significantly reduced his/her volume, which was very infrequent.

3.3.1.5 Measurements

The waveforms of all the target words were manually labeled in Praat. The pitch and intensity contour of the rhyme of each word was extracted using a Praat script. The script sampled the pitch and intensity from the beginning (t_0 = time at 0% of duration) of the rhyme to the end (t_{100} = time at 100% of duration), at intervals of 10 percent increments of the total duration. The purpose of this approach was to provide a normalized time frame within which to compare the pitch and intensity contours of different speakers. The script also measured the duration of the rhyme, which is simply the difference between t_{100} and t_0.

In this experiment, I *z*-normalized the F0 data points (F0z). This procedure helps to minimize the variation between subjects (Rose, 1993). The extracted pitch contour of a surface realization is represented by an ANOVA smoothing spline, giving the best fit across the normalized F0 (F0z) data points for all subjects and their 95% confidence interval (Gu, 2014). If two splines fall within each other's confidence envelope at a given point, their F0z at this particular point are not statistically different. Similarly, if there is no overlap between the two envelops, the F0z are statistically distinct from each other. I have also applied a similar statistical treatment to intensity data points, which are also z-normalized (Iz).

3.3.2 Results and discussion

I will discuss the results by addressing the research question raised at the beginning of this experiment. First, *what is the surface realization of tone in LT words and does it match the description laid out in previous sections?* Second, *is there any duration difference between different types of syllables and does it affect tonal patterns?* These questions will be answered in the sections below.

3.3.2.1 Surface realization of tones in LT words

In this section, I will present the surface realization of tones in monosyllabic, disyllabic, and polysyllabic words of LT. Relevant data is presented in this section, but detailed results of individual stimuli are presented in Appendix C.

Monosyllabic Words: The pitch contour of monosyllabic words is determined by the lexical tones /H, LH/, and their interaction with the onset consonant and the types of rhyme. The surface pitch contours of H and LH are given in Figure 3.11: [H] is realized as a falling pitch contour, and [LH] as a rising pitch contour. I attribute the falling pitch contour of the [H] to the presence of a L% boundary tone at the end of target words, which were uttered in isolation. This falling contour is more conspicuous in shorter words, which tend to have a brief plateau before falling. Based on these plots, it is immediately clear that the F0 onset of H is significantly higher than that of LH. Interestingly, in most cases, the F0 offset of H and LH are virtually the same.

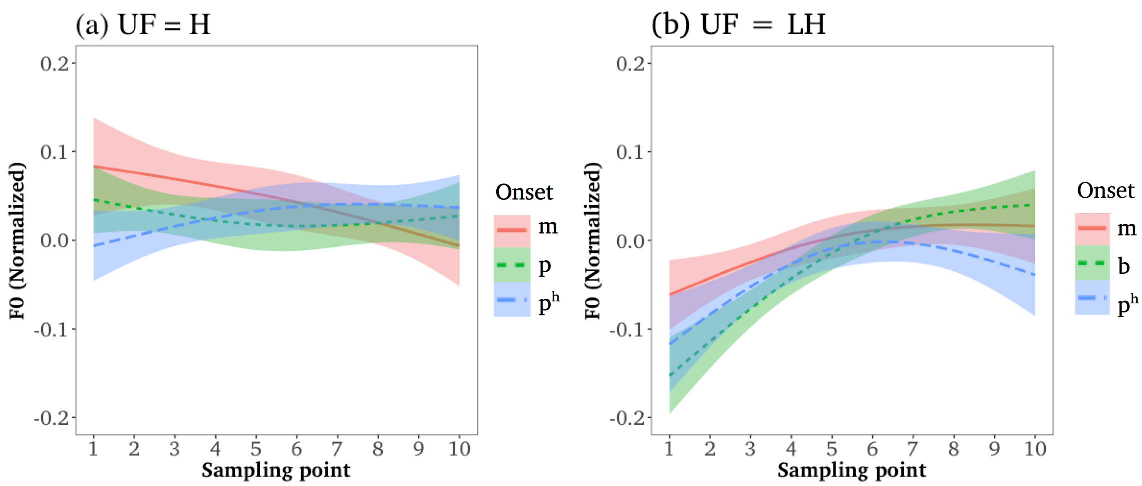


Figure 3.11: Surface realization according to onset of monosyllabic word.

Some intrinsic F0 effects can be attributed to onset consonants, as shown in Figure 3.11. /m/ raises the entire F0 contour of a tone, be it H or L. /p^h/ also seems to be associated to a slightly higher pitch than /p/ in syllables with a H tone, but this effect happens relatively late in the rhyme.

Another factor conditioning F0 contour variation in LT monosyllabic words is the interaction between the lexical tones (H/LH) and the rhyme types (CV/CV[-son] vs. CV:/CV[+son]). Words, with CV/CV? shape and H, tend to have a high falling pitch contour, and LH words of the same shape tend to have more level but still a moderately low-rising pitch contour. The pitch realization of LH provides supporting evidence that a syllable allows tonal crowding and can support up to two tones (cf. Monich, 2011). In general, short and long rhymes (μ vs $\mu\mu$) seem to have fairly similar tone contours, except the realization LH on CV: which has significantly lower onset and offset. This is further evidence that the syllable is the TBU instead of the mora.

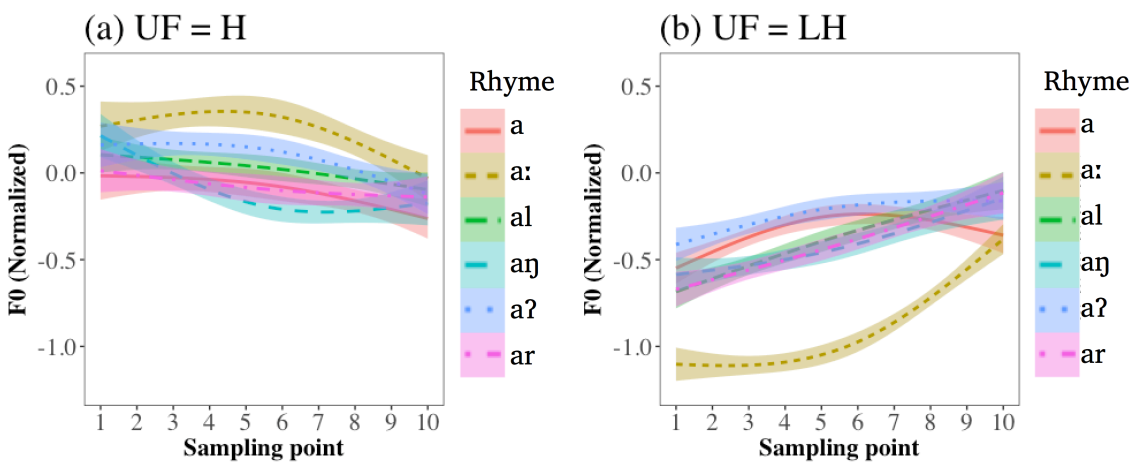
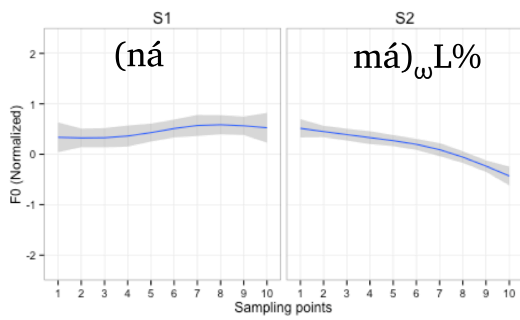


Figure 3.12: Surface realization according to rhyme of monosyllabic word.

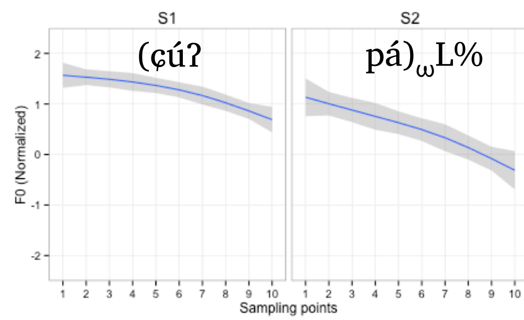
Disyllabic words: Based on my phonological analysis (see 3.2.3.1), there should be only two possible tone outputs in disyllabic words: [H] and [LH]. The pitch pattern of LT disyllabic words is, as expected, relatively straightforward. Disyllabic words that have an H lexical tone on their word-initial syllable (S1), such as *mna' ma* [námá] in Figure 3.13(a), are expected to have a phonological output of [H.H]. This is what we find in examples (a) and (c), where F0 is high and flat on the first syllable, maintains the same general height at the onset of the second syllable, and gently falls off the end of the second syllable because of the boundary tone L%. Example (b), *shug pa* [cuʔpa], has an high and falling initial syllable, which could be attributed to the presence of a glottal stop.

Example (d), *chu sel* [tɕʰù.ɕél] 'crystal', is more problematic: most subjects (six out of eight) produced it with a LH sequence instead of the expected [H] reported in earlier studies (Qu and Tan, 1983; Duanmu, 1992). At this point, I am unable to explain why there is inter-speaker variation in the tonal realization of this word,

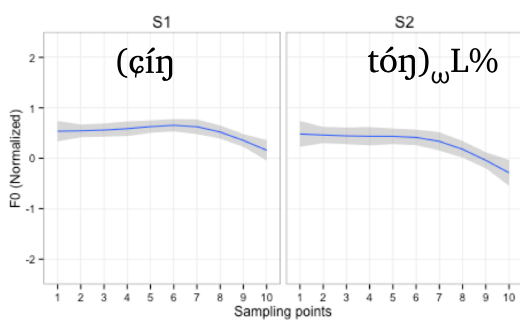
(a) Surface realization of *mna' ma* /na^{H}ma^{LH}/



(b) Surface realization of *shug pa* /ɕuʔ^{H}pa^{H}/



(c) Surface realization of *shing sdong* /ɕiŋ^{H}toŋ^{LH}/



(d) Surface realization of *chu shel* /tɕ^{h}u^{H}.ɕel^{H}/

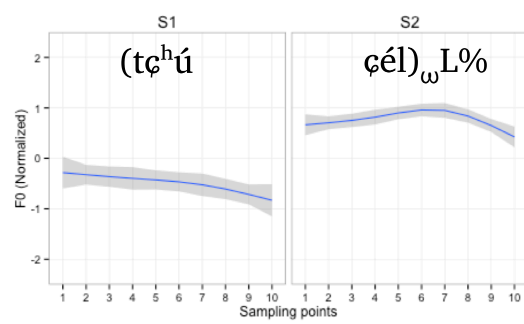


Figure 3.13: Examples of surface realization of [H] disyllabic words.

but what is clear is that the unexpected tone contour in Figure 3.13 is due to variation in its underlying tone rather than an aberrant phonetic realization of the H tone on a disyllable.

For the disyllabic words that have a LH lexical tone on their initial syllable (S1), such as the words shown in Figure 3.14, the expected phonological output is [L.H]. The L of the initial syllable has a low pitch onset and level pitch contour. This low level pitch contour is categorically different from the high pitch contour of H on the first syllable of the words in Figure 3.13. It is also much lower than the H of the second syllable. The phonological output for the word-final H of an underlyingly /LH.LH/ word has a surface realization that is similar to the /H.H/found in Figure 3.13. This suggests that the decomposed H of an underlying LH is identical

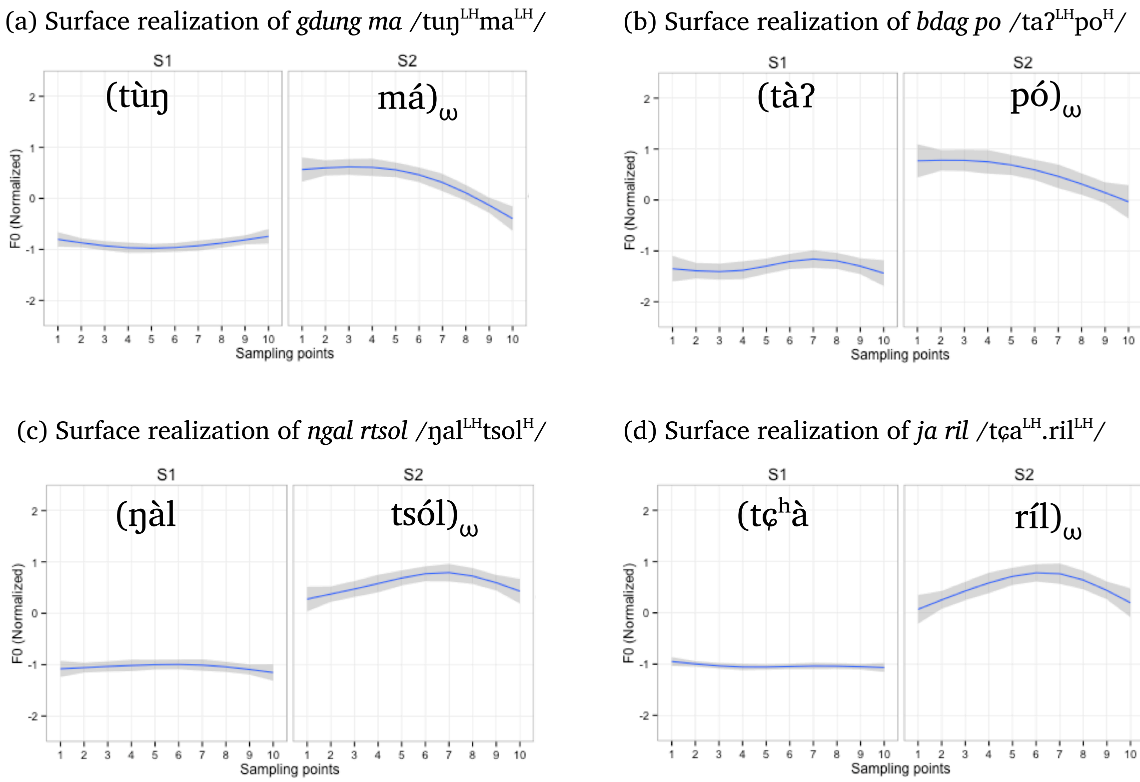


Figure 3.14: Examples of surface realization of [LH] disyllabic words.

to an underlying H, and that the LH contour is decomposed into two level tones, i.e., L and H.

Polysyllabic words: As mentioned in the discussion in §3.2, a polysyllabic word is considered to be a prosodic word group comprised of smaller prosodic words. The surface realization of polysyllabic, trisyllabic, and quadrisyllabic words is a concatenation of surface realizations of monosyllabic and disyllabic words. The examples that are presented below are just a confirmation of what was found in the monosyllabic and disyllabic words discussed earlier. The only additional elements are the Downstep rule already described in §3.2.3 and an anticipatory lowering observed in the H/L tones before a L tone.

First of all, I would like to present a pair of examples that illustrate that the

H of a non-decomposed LH does not downstep the following H, in contrast with the H of a decomposed LH. This is shown in Figure 3.15. The trisyllabic word, *ngo shes pa* /ŋo^{LH}ʃe^{H}pa^{L}/ 'acquaintance', has a phonological output of [LH][H][L], as shown in example (a). The LH is realized in the first syllable (S1) as a continuous rising contour and is followed by a high-falling contour on the second syllable, then a low plateau on the third syllable. The continuous rising pitch contour of the first syllable suggests that it is the surface realization of a non-decomposed [LH]. The F0 at the offset of the first syllable is virtually the same as the F0 at the onset of the second syllable, which suggests that there is no downstep on the H of the second syllable. The falling pitch contour is caused by an anticipatory lowering in anticipation of the L on the third syllable. In contrast, the second trisyllabic word, *gzhes ka rtse* /ʃi^{LH}ga^{T}tse^{H}/ 'Shigatse,' has a phonological output of [L.H][!H] as previously shown in example (b). The decomposed [L.H] is realized as two plateaux on the first and second syllable. The F0 offset of the H of the second syllable has significantly higher F0 than the onset of the following syllable, despite its H. This suggests that the H of the third syllable is downstepped. In short, these examples show that a non-decomposed [LH] does not downstep the following H, but that a decomposed [L.H] does.

(a) Surface realization of *ngo she's pa* /ŋo^{LH}ʃe^{H}pa^{H}/ (b) Surface realization of *gzhes ka rtse* /ʃi^{LH}ga^{T}tse^{H}/

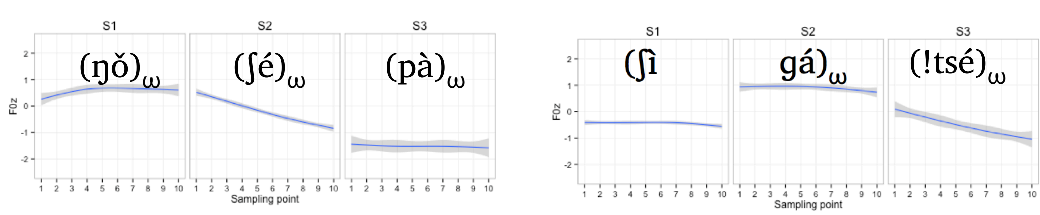


Figure 3.15: Surface realization of *ngo shes pa* /ŋo^{LH}ʃe^{H}pa^{L}/ 'acquaintance' and *gzhes ka rtse* /ʃi^{LH}ga^{T}tse^{H}/ 'Shigatse.'

I now turn to another example of downstep found in quadrisyllabic words.

This downstep is not captured in the earlier transcription provided by Qu and Tan (1983). For instance, *dkar ra dkar po* [ká.rá.kár.pó] 'white sugar,' in Figure 3.16 has an [H][!H] phonological output, where each H is linked to two syllables, as shown in the phonological derivation in §3.2.3. The F0 onset of the first H (kara) is significantly higher than that of the second H (karpo) as shown in Figure 3.16. It is also interesting to note that the F0 offset of the first H undergoes limited anticipatory lowering due to the following downstepped H.

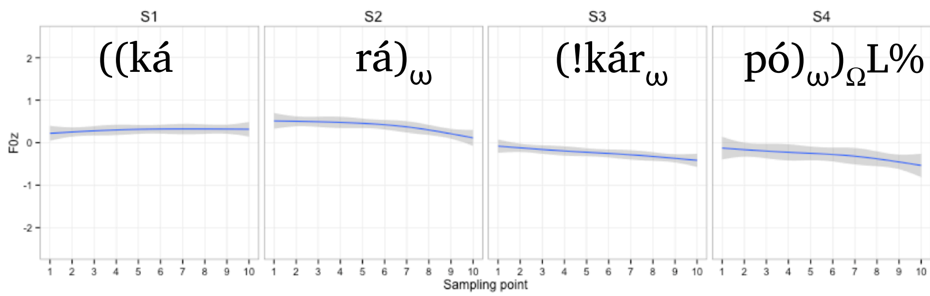


Figure 3.16: Surface realization of *dkar ra dkar po*/$ka^{H}ra^{H}!kar^{H}po^{H}$/ 'white sugar.'

In short, the pitch patterns observed in polysyllabic words are the result of the concatenation of the pitch patterns of monosyllabic and disyllabic prosodic words, with the addition of downstep.

3.4 Experiment II: Status of word stress of LT words

In this experiment, I will try to address the following research question: *Is there any stress pattern in the polysyllabic words of LT?*

I do this by looking at quadrisyllabic words, that are the most likely to exhibit stress independently form edge effects. I focused on three acoustic parameters: rhyme duration, pitch, and vocal intensity. These acoustic cues are usually associated with stress, cross-linguistically. I would like to remind the reader that both rhyme duration and pitch are used as cues for other phonological contrasts in LT: the pitch is the main cue for tonal contrast, whereas rhyme duration is the main cue for vowel length in an open syllable (i.e., V vs. V:).

3.4.1 Methods

The basic methodology for Experiment II is identical to that used for Experiment I (see §3.3.1). The speech samples were collected from the same participants with the same recording technique. Only a small subset of the stimuli from Experiment I was selected for analysis in Experiment II, and a different aggregation of the data was employed. These two aspects of the methodology are discussed in the following subsection.

3.4.1.1 Stimuli

Polysyllabic words were chosen for Experiment II in order to control for known cross-linguistic characteristics of stress. These four main characteristics are culminative stress (one maximally prominent peak), demarcative stress (marking morphological boundaries), rhythm (well-defined strong and weak syllables), and quantity sensitivity (syllable weight) (Kager, 2006, p.196-198). All of these characteristics can be identified effectively in polysyllabic words if they exist.

The following six quadrisyllabic words in Table 3.13 were selected as stimuli for Experiment II. All selected quadrisyllabic words are prosodic word groups that consist of a pair of disyllabic prosodic words. These prosodic words only have Hs in their phonological output. They contain a variety of syllable weights and rhyme duration, e.g., CV, CVO, CVN[9]. Following previous work (cf. Hayes, 1976), stress, if present, is expected to surface on either the first or second syllable of each prosodic word, and there could be additional prominence at the prosodic word group level (see §3.2.2.4).

WT	Segments	Gloss
dka' las khag po	((ká.lék)$_\omega$(kʰáʔpó)$_\omega$)$_\Omega$	'difficult'
dkar ra dkar po	((ká.rá)$_\omega$(kárpó)$_\omega$)$_\Omega$	'white sugar'
dkar po dkar rkyang	((kárpó)$_\omega$(kárkʲáŋ)$_\omega$)$_\Omega$	'pure white'
skye ba snga ma	((kéwá)$_\omega$(ŋámá)$_\omega$)$_\Omega$	'former life'
pha ma gcig pa	((pʰámá)$_\omega$(tɕíʔpá)$_\omega$)$_\Omega$	'same parent'
pu pu khu shud	((púpú)$_\omega$(kʰúʃýʔ)$_\omega$)$_\Omega$	'name of a Tibetan bird'

Table 3.13: Quadrisyllabic words with all H disyllabic prosodic words.

3.4.1.2 Measurement

If there is stress in LT, stressed and unstressed syllables should exhibit significant asymmetries in pitch, vocal intensity, and rhyme duration. The maximum pitch, average intensity, and rhyme duration of each syllable were extracted from the recorded sound file using Praat. Values of maximum pitch, intensity, and duration were normalized to minimize the idiosyncratic characteristics of individual participants.

[9]Unfortunately, there are no suitable all-H words that also have a long vowel (V:) syllable in my current data set.

3.4.2 Results and Discussion

The data were analyzed using linear mixed models implemented in the $lmerTest$ package of R. Three models were built with the three dependent variables discussed in the previous section (maximum normalized F0, average normalized intensity, normalized rhyme duration). The fixed and random effects incorporated in the initial linear models, along with their levels are presented in Table 3.14.

Fixed effects	Levels	Remarks
Syllable position	S1, S2, S3, S4	
Token	T1, T2, T3	Each word is repeated three times during elicitation, i.e. three tokens
Duration	Normalized rhyme duration (dz)	Not included in rhyme duration model
All relevant interactions		
Random effects	Levels	
Participants	M1,M2,M3,M4,M5,M6,M7,F1	
Words	As listed in Table 3.13	

Table 3.14: The fixed and random effects of linear mixed model used in Experiment II.

The fixed effects included in the models are Syllable position, Token, and Rhyme Duration, along with all their interactions. Rhyme Duration was excluded as a fixed effect in the model in which it is the dependent variable. Participants and words were the random effects that were included in each of these models. I maximized the random effect structure by including intercepts for both random effects, as well as by-participant and by-word random slopes for all models, as recommended by Barr et al. (2013). Results for the random slopes will not be reported here, as they only capture limited variance.

Duration of rhyme: As shown in Table 3.15, the variance of the random effect Word is much greater than that of Participants. This is expected since the data was z-normalized by speaker. The residual is relatively high, which is likely to reflect variation in speech rate across the wordlist.

Groups	Name	Variance	Std.Dev.
Participant	(Intercept)	0.001473	0.03838
Word	(Intercept)	0.126707	0.35596
Residual		0.393314	0.62715

Table 3.15: Estimates of random effects on normalized duration of rhyme of quadrisyllabic words.

According to the estimates of Table 3.16, the final syllable position (S4) has the most significant fixed effect. I attribute this to phrase-final lengthening. Similarly, the final token of an elicitation phrase (T3) has a longer rhyme duration, which is also attributed to utterance-final lengthening. As there is no significant rhyme duration difference between the first and second syllable of the first prosodic word, there is no evidence that duration is utilized as an acoustic cue of stress.

	duration∼ Syllable×Token + (1|Participant) + (1|Word)					
	Estimate	Std. Error	df	t value	Pr(>|t|)	p
(Intercept)	-0.33	0.17	9.14	-1.90	0.09	
SyllableS2	0.09	0.13	557.87	0.67	0.50	
SyllableS3	-0.12	0.13	557.73	-0.92	0.36	
SyllableS4	**0.48**	**0.13**	**557.73**	**3.80**	**0.00**	***
TokenT2	-0.24	0.13	557.73	-1.89	0.06	
TokenT3	-0.28	0.13	557.90	-2.16	0.03	*
SyllableS2:TokenT2	0.14	0.18	557.74	0.76	0.45	
SyllableS3:TokenT2	0.18	0.18	557.66	1.03	0.31	
SyllableS4:TokenT2	0.12	0.18	557.66	0.69	0.49	
SyllableS2:TokenT3	0.09	0.18	557.74	0.51	0.61	
SyllableS3:TokenT3	0.23	0.18	557.66	1.27	0.21	
SyllableS4:TokenT3	0.11	0.18	557.66	0.63	0.53	

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Table 3.16: Estimates of fixed effects on normalized duration of rhyme of quadrisyllabic words. Bold marks significant fixed effects.

F0: According to the estimates in Table 3.17, the random effects of participants and words are relatively small.

Groups	Name	Variance	Std.Dev.
Participant	(Intercept)	0.03101	0.1761
Word	(Intercept)	0.08002	0.2829
Residual		0.24198	0.4919

Table 3.17: Estimates of random effects on maximum normalized pitch of a syllable of quadrisyllabic words.

Table 3.18 reports the fixed effects. In Table 3.18, the two most significant fixed effects ($p < 0.001$) are the third and fourth syllable positions of the quadrisyllabic words. The third syllable is situated at the boundary between two prosodic words, and the fourth syllable is at the word/phrase-final position. This lower maximum F0 found in both syllables compared to the first syllable can be attributed to downstep, which I claim is a phonological encoding of the prosodic word boundary within a prosodic word group (cf. §3.2.3.2). This downstep is different from demarcative stress, because stress usually increases prominence and only affects one syllable, not two. Furthermore, there is no significant maximum F0 difference between the first and second syllable, or between the third and fourth syllable. The absence of difference between the first (S1) and second (S2) syllable is shown in Table 3.18, while the absence of a difference between the third (S3) and fourth (S4) syllable is obtained by refitting the same model with S3 as the reference level for the Syllable factor. When we do this, the results for S4 (compared with S3) are as follows: Estimate = -0.040; Std Error = 0.116; df = 530.10; t-value = -0.347; Pr(> |t|) = 0.728. Thus, there is no evidence that F0 is utilized as an acoustic cue of stress in LT.

The fact that there is a significant interaction between duration and the last two syllables (S3 and S4) of a quadrisyllabic word is unexpected. Corresponding estimates suggest that when the duration of these syllables increases, F0 goes up slightly. This seems to correspond to a less pronounced effect of downstep in very

F0zMax∼ Syllable×Token×duration + (1|Participant) + (1|Word)						
	Estimate	Std. Error	df	t value	Pr(>|t|)	p
(Intercept)	0.89	0.15	13.54	5.77	0.00	***
SyllableS2	0.15	0.11	529.48	1.36	0.17	
SyllableS3	**-0.41**	**0.12**	**529.38**	**-3.50**	**0.00**	***
SyllableS4	**-0.45**	**0.11**	**529.58**	**-4.09**	**0.00**	***
TokenT2	-0.08	0.13	529.56	-0.58	0.56	
TokenT3	-0.20	0.14	530.00	-1.42	0.16	
dz	-0.05	0.11	532.57	-0.47	0.64	
SyllableS2:TokenT2	-0.07	0.18	529.31	-0.40	0.69	
SyllableS3:TokenT2	-0.03	0.18	529.33	-0.14	0.89	
SyllableS4:TokenT2	0.12	0.17	529.35	0.74	0.46	
SyllableS2:TokenT3	-0.09	0.19	529.41	-0.48	0.63	
SyllableS3:TokenT3	-0.15	0.19	529.24	-0.77	0.44	
SyllableS4:TokenT3	-0.19	0.18	529.66	-1.06	0.29	
SyllableS2:dz	0.09	0.15	532.12	0.57	0.57	
SyllableS3:dz	**0.33**	**0.16**	**530.11**	**2.10**	**0.04**	*
SyllableS4:dz	**0.32**	**0.13**	**531.26**	**2.53**	**0.01**	*
TokenT2:dz	0.26	0.17	529.50	1.52	0.13	
TokenT3:dz	0.34	0.19	529.98	1.79	0.07	
SyllableS2:TokenT2:dz	-0.10	0.25	529.40	-0.38	0.70	
SyllableS3:TokenT2:dz	-0.29	0.24	529.43	-1.20	0.23	
SyllableS4:TokenT2:dz	-0.23	0.20	529.27	-1.19	0.23	
SyllableS2:TokenT3:dz	-0.05	0.25	529.19	-0.21	0.83	
SyllableS3:TokenT3:dz	-0.43	0.26	529.31	-1.70	0.09	
SyllableS4:TokenT3:dz	-0.39	0.22	529.85	-1.79	0.07	

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Table 3.18: Estimates of fixed effects on maximum normalized pitch of a syllable in quadrisyllabic words. Bold marks significant fixed effects.

slow speech rates, which could result from an overly careful elicitation style[10].

Vocal intensity: According to Table 3.19, the variance accounted for by the random factors participants and words is insignificant. The relatively large residual is likely due to variation in loudness across the word list.

[10]According to the results in Table 3.18, there are no statistically detectable F0 effects which are attributed to the sequence of tokens (i.e. T1, T2, T3).

Groups	Name	Variance	Std.Dev.
Participant	(Intercept)	0.02388	0.1545
Word	(Intercept)	0.02176	0.1475
Residual		0.37084	0.6090

Table 3.19: Estimates of random effects on normalized Intensity of a syllable of quadrisyllabic words.

$Iz_{avg} \sim$ Syllable×Token×duration + (1|Participant) + (1|Word)						
	Estimate	Std. Error	df	t value	Pr(>|t|)	p
(Intercept)	1.03	0.13	46.38	8.03	0.00	***
SyllableS2	**-0.44**	**0.14**	**547.04**	**-3.23**	**0.00**	**
SyllableS3	**-0.60**	**0.15**	**546.63**	**-4.15**	**0.00**	***
SyllableS4	**-1.28**	**0.13**	**547.48**	**-9.68**	**0.00**	***
TokenT2	-0.24	0.16	546.90	-1.50	0.13	
TokenT3	**-0.47**	**0.18**	**547.80**	**-2.70**	**0.01**	**
dz	0.05	0.14	552.75	0.36	0.72	
SyllableS2:TokenT2	0.13	0.22	546.44	0.60	0.55	
SyllableS3:TokenT2	0.17	0.23	546.15	0.77	0.44	
SyllableS4:TokenT2	0.10	0.20	546.14	0.49	0.62	
SyllableS2:TokenT3	0.23	0.23	546.56	0.99	0.32	
SyllableS3:TokenT3	0.15	0.24	545.97	0.63	0.53	
SyllableS4:TokenT3	0.07	0.21	546.70	0.31	0.76	
SyllableS2:dz	-0.17	0.18	552.23	-0.92	0.36	
SyllableS3:dz	0.31	0.19	548.17	1.63	0.10	
SyllableS4:dz	-0.03	0.16	549.83	-0.20	0.84	
TokenT2:dz	0.22	0.21	546.46	1.05	0.30	
TokenT3:dz	0.01	0.23	547.27	0.05	0.96	
SyllableS2:TokenT2:dz	0.19	0.31	546.58	0.60	0.55	
SyllableS3:TokenT2:dz	-0.02	0.30	546.24	-0.08	0.93	
SyllableS4:TokenT2:dz	-0.30	0.24	545.95	-1.23	0.22	
SyllableS2:TokenT3:dz	0.26	0.31	545.97	0.82	0.41	
SyllableS3:TokenT3:dz	0.01	0.32	546.02	0.02	0.98	
SyllableS4:TokenT3:dz	-0.28	0.26	547.47	-1.04	0.30	

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Table 3.20: Estimates of fixed effects on average normalized intensity of a syllable of quadrisyllabic words. Bold marks significant fixed effects.

According to Table 3.20, the first syllable of quadrisyllabic word has the highest vocal intensity. All subsequent syllables undergo significant reduction in intensity. The progressive lowering of vocal intensity across a polysyllabic word is expected because of the decreasing subglottal pressure. Similarly, the last token of a sequence of three tokens also has significant reduction in vocal intensity, probably for the same reason. In essence, vocal intensity is not used as an acoustic cue for stress in LT.

Based on this experiment, I conclude that there is no evidence to support the proposal that LT has a stress system realized through rhyme duration, pitch, or vocal intensity. These acoustic cues, however, are used to signal prosodic word boundaries and word-final syllables. In short, there is no evidence for overt stress, as one can account for tonal processes by formalizing the low tone as LH rather than invoking a stress-assigned H. It is more economical to analyze LT as a stressless language.

In this chapter, I have established that LT has no word stress, and that its tonal pattern can be accounted for by invoking three simple lexical tones (H, LH and L), two prosodic constituents (the prosodic word and the prosodic word group) and a small number of tone rules. In the next chapter, I show how word tone interacts with larger intonational units.

Chapter 4

The intonational phonology of LT

Before opening this chapter on LT intonation, it is important to point out that the communicative functions of LT are mainly realized through morphosyntactic strategies, which include suppletion of verb stems and use of discourse particles. For example, some verb stems in LT preserve the Classical Tibetan suppletive imperative, as shown in Table (4.1).

Present-future	**Past**	**Imperative**	**Gloss**
’gro drɔ̌/drʊ̌	*phyin/song* tɕʰíː	*song* gʲǔ	‘to go’
’ong jɔ̌̃ː/jʊ̌ŋ	*’ongs* jɔ̌̃ː	*shog* ɕɔ́ː	‘to come’
byed/bya chěʔ/chǎʔ	*byas* chɛ̌ʔ	*byos* chø̌ʔ/chěʔ	‘to do’
rgyag/brgyag kʲǎʔ	*bragyab* kʲǎp	*rgyob* kʲǒp/kʲǎp	‘to construct’
lta/blta tá	*ltasb* tɛ́ʔ	*ltos* tǿ	‘to look’
za/bza’ sǎ	*bzas* sɛ̌ʔ	*zo* sǒ	‘to eat’

Table 4.1: Suppletion of verb stems from Tournadre (1996, p.418-419) and Denwood (1999, p.108).

Discourse markers are the most prevalent way of expressing communicative functions in LT. Like many other East Asian languages, LT uses final particles to express the interrogative (Q)[1] or the imperative (IMP), as illustrated in Examples (11), and (12), respectively. By combining various particles and forms of verb stem, LT can also communicate polite imperative, optative imperative, and friendly imperative (Denwood, 1999, p.167-168). Interestingly, these imperatives have the same intonational pattern (i.e., phrase-final falling) as declarative statements.

(11) *su yin pa* [sú jǐn -pèː] (Tournadre and Dorje, 2003, p.86)
you are-Q
'(Who) are you?'

(12) *gzigs rogs gnang* [sǐ -ronaŋ](Tournadre and Dorje, 2003, p.196)
go IMP
'Look!'

The remainder of this chapter focuses on the limited use of intonational patterns for communicative functions in LT. This includes a description of intonational phrases (§4.1), communicative functions and boundary tones (§4.2), information structure (§4.3), and a brief discussion of the unique spelling chant melody (§4.4). At the end of this chapter, I provide a concluding summary of the tonal inventory of LT (§4.5).

4.1 The intonational phrase and its cues

The melodic contour of an LT sentence is achieved by combining the word-level tonology discussed in the previous chapter with the boundary tones of intonational phrases and the melodic cues marking information structure. Thus, LT can be classified as a head/edge prominence language (a hybrid prominence system) instead of a solely head-based or edge-based prominence language (c.f. Jun, 2014).

[1] It is important to point out that the tone of the particle is not necessarily the boundary tone associated with the communicative function. For instance, the interrogative particle has a low tone, but the interrogative intonational pattern has a rising pitch contour.

…ι…	Intonational phrase
|	
…Ω…	Prosodic word group
|	
…ω…	Prosodic word
|	
…σ…	Syllable
|	
…μ…	Mora

Figure 4.1: The prosodic structure of LT.

The intonational phrase is the largest constituent of the LT prosodic structure, as shown in Figure 4.1. Syntactically, it usually matches clauses (this includes root clauses, syntactic adjuncts, parentheticals and more). Prosodically, it usually consists of one or more prosodic word groups (Ω). A new intonational phrase boundary can also be triggered by discursive factors such as hesitations and corrections. Occasionally, LT intonational phrases are marked by a discourse particle at the phrase-initial position, such as *ání/áné* 'then or and'. This, however, is an idiosyncratic feature of some speakers during spontaneous speech. The boundaries of intonational phrases have three main phonetic cues: *a*) *pitch level-reset* —declination is partially reset at the beginning of intonational phrases, *b*) *pause* —there is sometimes a pause between two intonational phrases; *c*) *phrase-final lengthening* —the duration of the final syllable of an intonational phrase is significantly longer than that of preceding syllables; *d*) *boundary tone* —the last syllable of the intonational phrase can bear a boundary tone. The latter two cues are shown in Figure 4.2. It is important to point out that current research has found no evidence in favour of the existence of an utterance phrase or of large recursive intonational phrases. An utterance is merely a concatenation of intonational phrases, as in Figure 4.3.

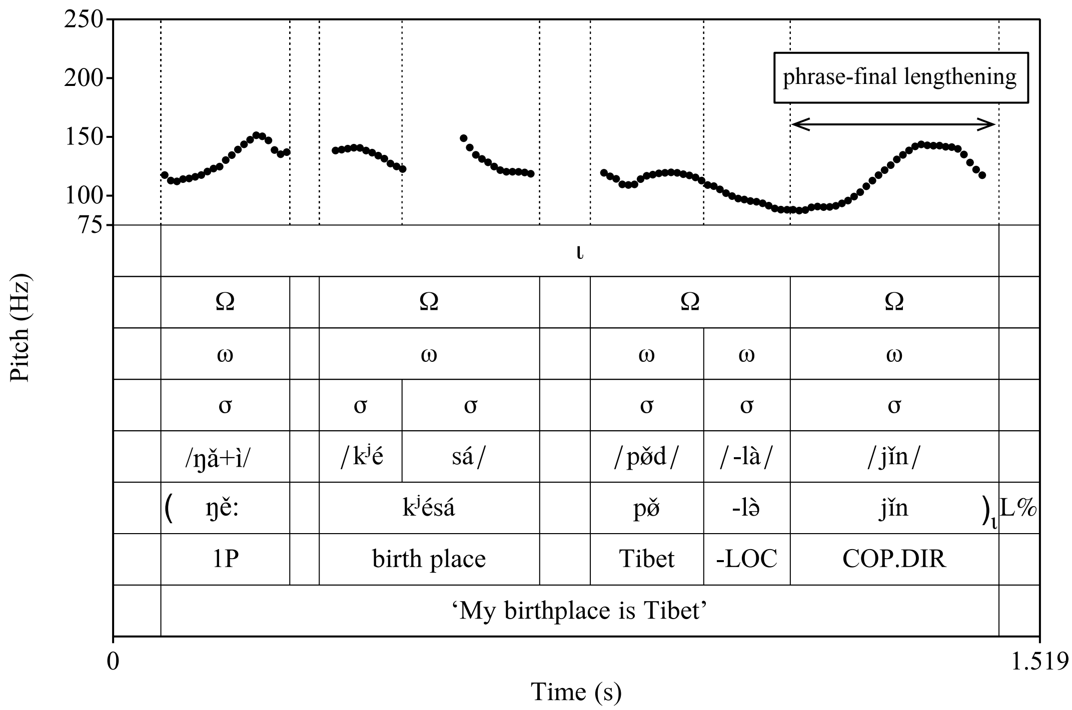


Figure 4.2: The intonational contour of a declarative intonational phrase with a phrase-final L%: *nga skyes sa bod las yin* /ŋa^{H} kʲesa^{LH} pø^{LH}-la^{L} jin^{LH}/ → [ŋǎ kʲésá pø̌d-là jǐn] ‘My birthplace is Tibet’

$$\begin{matrix} X\% & X\% & X\% \\ (\ldots)_\iota & (\ldots)_\iota & (\ldots)_\iota \end{matrix}$$

Figure 4.3: Concatenation of intonational phrases.

4.2 Communicative functions and boundary tones

As mentioned earlier, boundary tones attach to the end of intonational phrases. There are only two types of phrase-final boundary tones in LT: L% and H%. The surface realizations of these phrase-final boundary tones agree with the *Frequency Code*, where L% communicates finality (e.g., declarative sentence) and H% communicates non-finality (e.g., interrogative sentence) (Ohala, 1984). Boundary tones are associated to the last syllable (i.e., TBU) of the intonational phrase. As L% and H% are respectively lower and higher than corresponding lexical tones, the phrase-final L% and the tone (H or L) of the last syllable form a falling contour (i.e., H + L% or L + L%) , while the phrase-final H% and the tone of the last syllable form a rising contour (i.e., H + H% or L + H%).

In the following sections, I will present the major communicative functions associated with boundary tones in LT intonation, which include continuatives (§4.2.1), non-continuatives (§4.2.2), and interrogatives (§4.2.3)

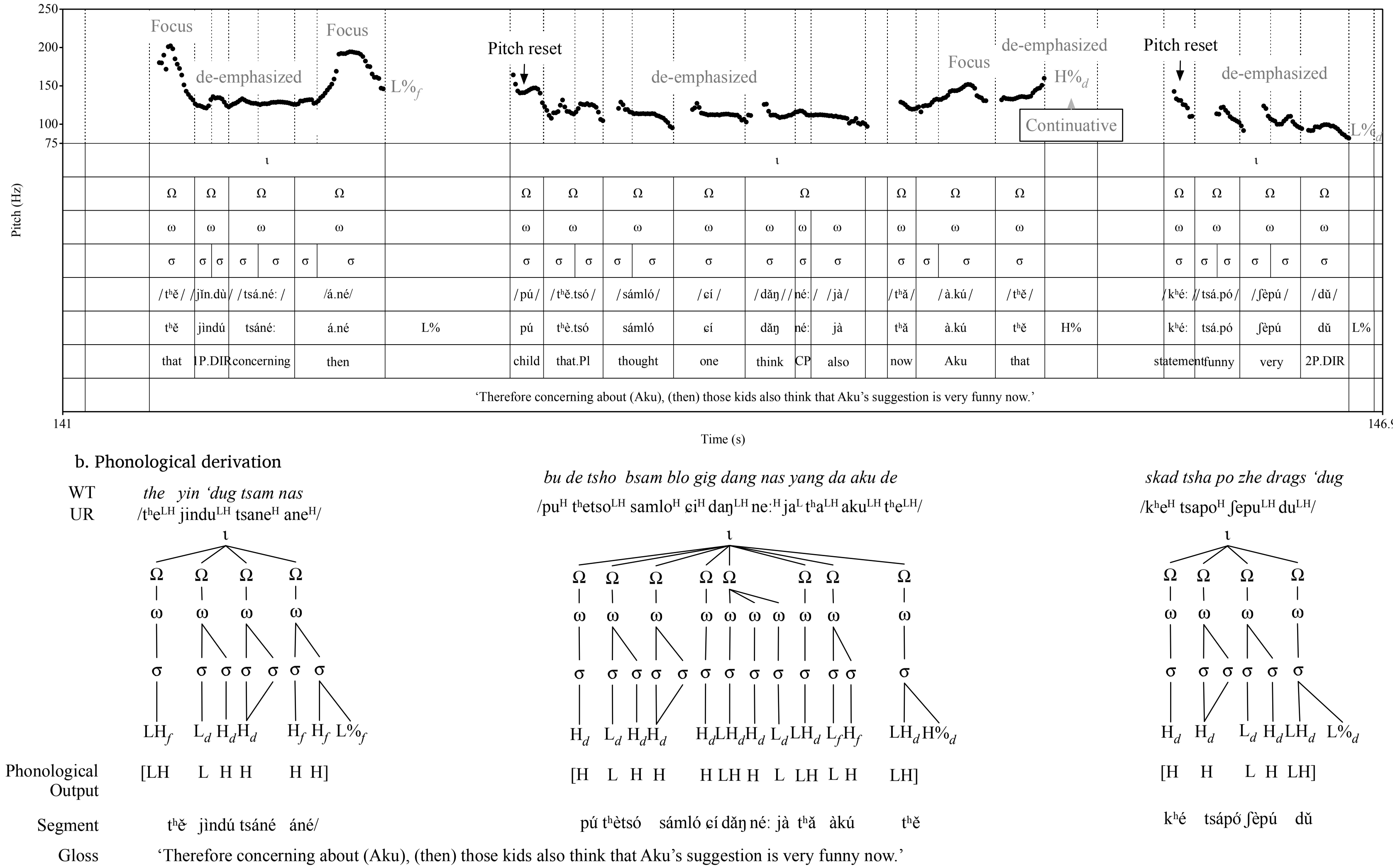


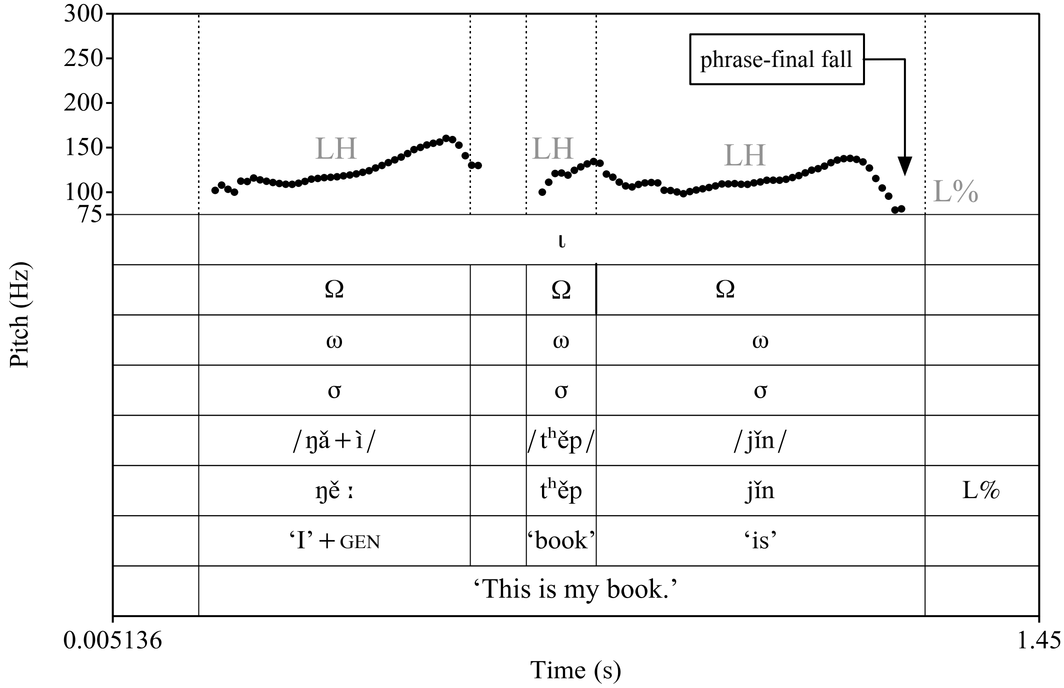
Figure 4.4: An example of continuative clauses extracted from a monologue.

4.2.1 Continuative intonation

Continuative clauses end on a H% boundary tone which signals that additional information has to be expected. An example of this intonation pattern is shown in Figure 4.4. Figure 4.4 contains three intonational phrases. The first two are continuatives. The first intonational phrase ends in the particle *nas* [ánέ], which is a morphosyntactic strategy for marking continuation. The second intonational phrase, however, ends in H%, which is an intonational strategy for marking continuation. The second and third intonational phrases also start with a pitch reset, as most intonational phrases in LT, but in the current examples, pitch reset is dampened by post-focal de-emphasis. Focus and de-emphasis will be fully addressed in §4.3.

In Figure 4.4 also contains examples of focused constituents in the first and second intonational phrases which tones are marked with a subscripted "f". Pitch-level reset at the phrase-initial boundaries of second and third intonational phrases, but in the current example they are dampened by the post-focal de-emphasis. These de-emphasized tones are marked with a subscripted "d" (cf. §4.3.) The interaction between tones (lexical and boundary) and information structure will be discussed in detail in §4.3.

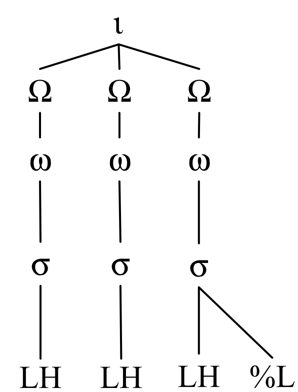
a. Intonation pattern



b. Phonological derivation

WT *nga ki deb yin*

UR /ŋa^{LH}+i^{L}/+tʰep^{LH}+jin^{LH}/



PO [LH LH LH]

Segment ŋě: tʰěp jǐn

Gloss 'This is my book.'

Figure 4.5: An example of a declarative statement: *nga gis deb yin* /ŋeː^{LH} + tʰep^{LH} + jin^{LH}/ '(The) book is mine.'

4.2.2 Non-continuative intonation

Non-continuative intonation is normally marked on the last intonational phrase of an utterance. It is marked with a phrase-final L%, which represents phrase-final lowering. The phrase-final lowering is consistent with the crosslinguistic observation that declarative sentences usually have a "terminal fall" (Cruttenden, 1997; Hirst and de Cristo, 1998).

An example of non-continuative intonation extracted from a monologue is given in Figure 4.5, where the last LH tone is followed by a dramatic fall caused by the L%.

The second example of non-continuative intonational pattern in a declarative statement is shown in Figure 4.6. This declarative statement consists of three intonational phrases. Each intonational phrase is marked by a phrase-final L%. That the first two intonational phrases are non-continuative, against the apparent semantic context of the sentence, is due to the fact that the participant hesitated after the first intonational phrase and paused after the second one, before reconsidering his decision and continuing the sentence. The pitch reset at the beginning of the third intonational phrase allows the tones to start at a similar F0 level as in the previous intonational phrase despite the preceding phrase-final fall. Note that the final syllable of example in Figure 4.6 is under focus, which has the effect of expanding its F0 range. I will come back to such focal effects in §4.3.

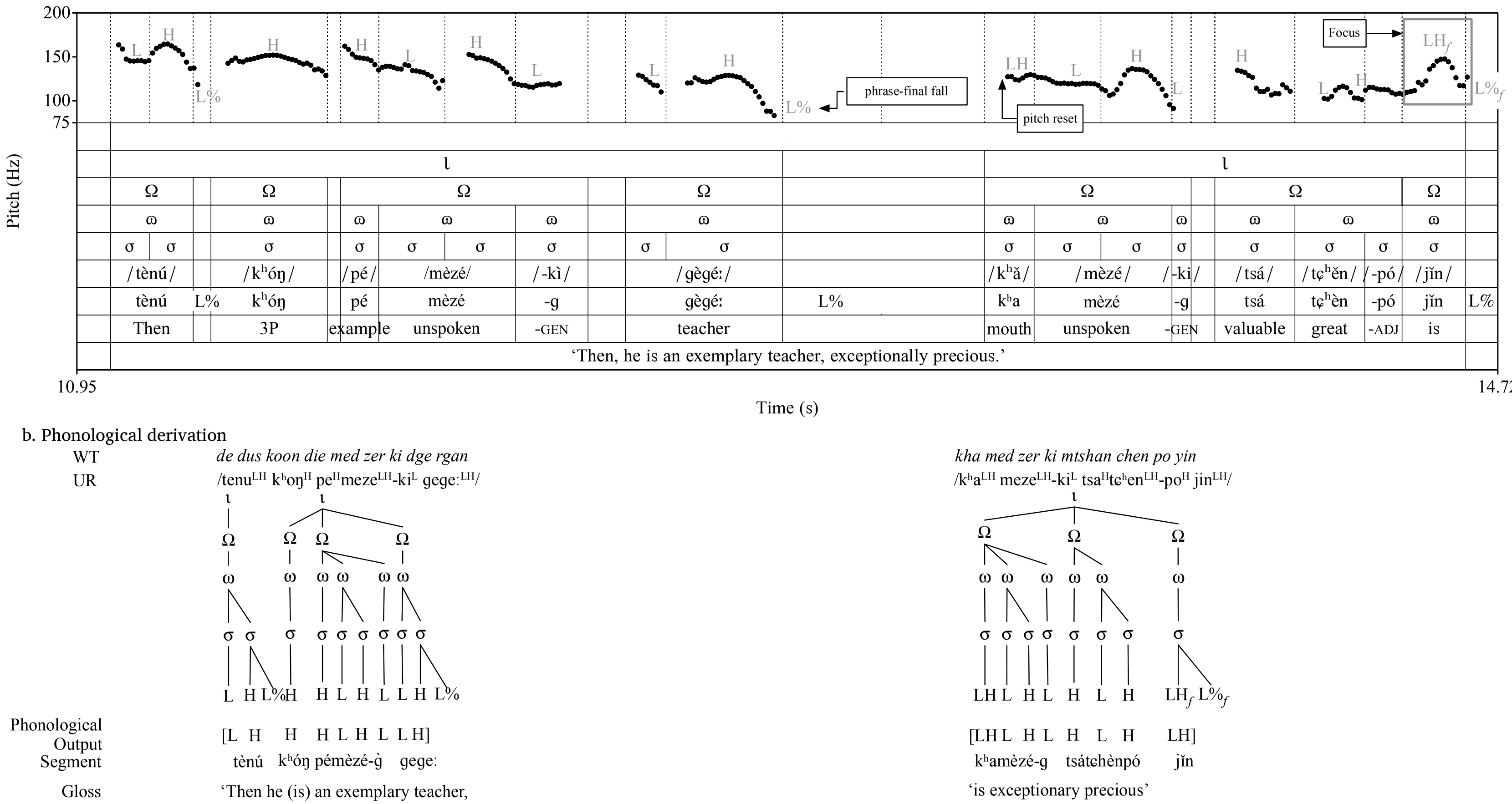
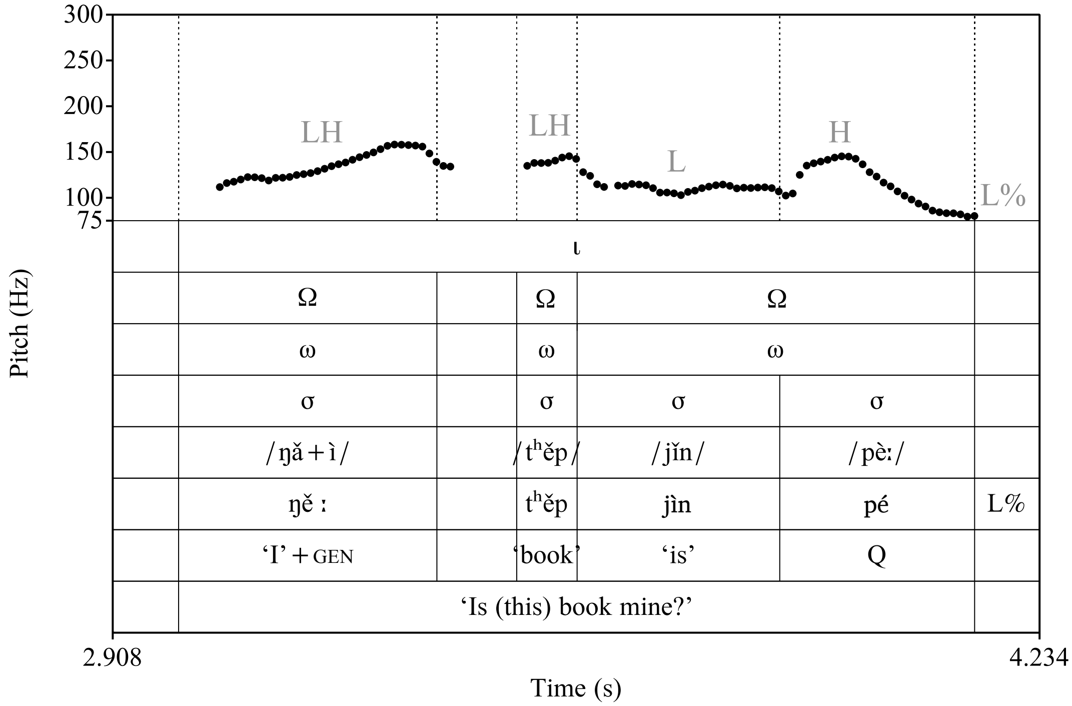


Figure 4.6: An example of a declarative statement extracted from a monologue.

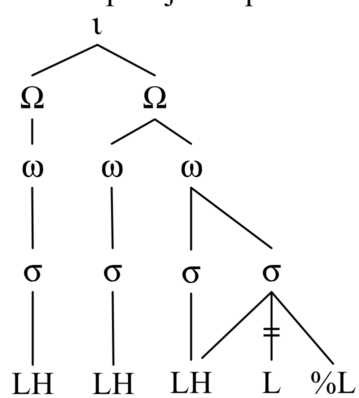
a. Intonation pattern



b. Phonological derivation

WT *nga ki deb yin pas*

UR /ŋa^{LH}+i^{L}/+t^{h}ep^{LH}+jin^{LH}+pe:^{L}/



PO [LH LH L H]

Segment ŋě: tʰěp jin pé:

Gloss 'Is (this) book mine ?'

Figure 4.7: An example of an interrogative statement with interrogative particle: *nga gis deb yin pa* /ŋe:^{LH} + t^{h}ep^{LH} + jin^{LH} + pe:^{L}/ 'Is (this) my book?'

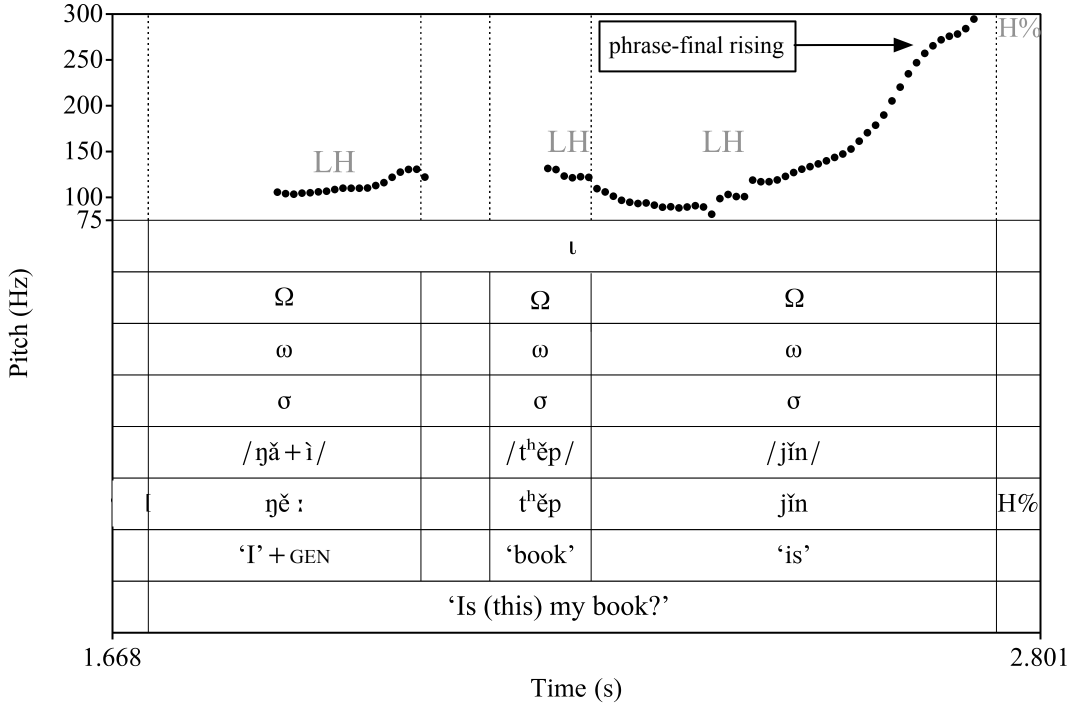
4.2.3 Interrogatives

There are three means of forming an interrogative clause in LT: *a*) with an interrogative form of the tense/aspect/mode (TAM) suffixes (Tournadre and Dorje, 2003, p.464); *b*) with an interrogative particle *pa/pas* /-pa/pe:/ (Tournadre and Dorje, 2003, p.85-86); and *c*) with a phrase-final H%. The interrogative form of TAM suffixes is mainly used in *Wh*-questions, where the *Wh*- word is in-situ, and the interrogative form of the TAM suffix is used with the verb stem. The interrogative particle, on the other hand, is mainly used in *Yes/No*-questions. The interrogative particle *pas* /-pè:/ is usually attached to the end of a TAM suffix or copula[2]. The general intonation pattern of these two types of interrogative statements is similar to that of a non-continuative clause, which is marked with a phrase-final L%, as illustrated in Figure 4.7.

When there is no interrogative particle, the L% is replaced with a H%, which also conveys of sense of surprise. This yields a phrase-final rising pitch contour instead of a phrase-final falling contour. If the phrase-final word tone is an H, the adjacent H% significantly increases the pitch offset of the last syllable. A Similar strategy is used in Japanese and Bantik Indonesian (Utsumi, 2016). An example of this intonational pattern and its phonological derivation are presented in Figure 4.8.

[2]The onset stop of the interrogative particle will assimilate to the coda velar stops, yielding the following variants: *pas* /-pè:/→[ngè:]/ng___ or *pas* /pè:/→[kè:]/k___.

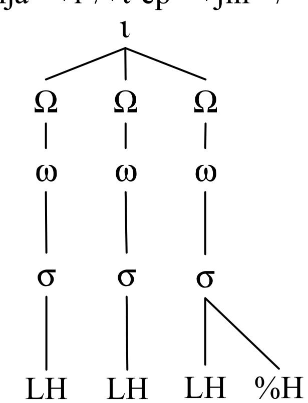
a. Intonation pattern



b. Phonological derivation

WT *nga ki deb yin*

UR /ŋa^{LH}+i^{L}/+t^{h}ep^{LH}+jin^{LH}/



PO [LH LH LH]

Segment ŋě: tʰěp jǐn

Gloss 'Is (this) my book?'

Figure 4.8: An example of an interrogative statement with interrogative intonation: *nga gis deb yin* /ŋe:^{LH} + t^{h}ep^{LH} + jin^{LH}/ 'This is my book?'

4.3 Information structure

Lambrecht (1994, p.5) defines information structure as the "component of sentence grammar in which propositions as a conceptual representation of states of affairs are paired with lexico-grammatical structures in accordance with the mental states of interlocutors who use and interpret these structures as units of information in given discourse contexts." LT defines the basic information structure by heavily relying on emphatic particles and syntactic structures instead of intonation. This characteristic is not uncommon in Sino-Tibetan tonal languages (Michaud and Brunelle, 2016).

One example of syntactic marking of information structure is the emphatic maker, *nid* [ɲìː] 'that very,' is simply attached to the entity, as shown in Example (13), to identify this specific entity *rgyal po* [gʲàlpó] 'king' to the listener.

(13) *rgyal-po nid* [gʲàlpó ɲìː]
king that very
'That very king.' (Garrett and Hill, 2015, p.36)

An example of morphological marking of information structure is the split-intransitive case-marking system where subjects of transitive and unergative verbs optionally receive ergative marking (DeLancey, 1990, p. 289). According to Tournadre (1991, p.105), the ergative argument has three "grades" that serve to mark the information status of a particular subject: *a*) zero anaphora (being the topic, the agent is implied but not formally mentioned, see (14); *b*) the agent is marked with the absolutive (neutral)[3]; and *c*) the agent is marked ergative (emphasis, see (16)). In example (16), the subject is marked with ergative and has contrastive emphasis.

[3] I have altered the original term which is 'rhetorically flat', see (15).

(14) *∅* *deb* *bklogs ki ’dug* [tép lɔ̀.gíduʔ]
ZERO book.ABS read.TAM:IPAST.IDIR
‘he/she is reading a book’

(15) *da* *nga* *gro-gi.yin* [dǎ ŋǎ drǒ-kì-jǐn]
now 1P+ABS go.PRES-FUT.EGO.VOL
‘Now, I’ll go.’ (Tournadre, 1991, p.101)

(16) *khyed zhugs* *a* *nga-s* *phyin-dgos* [kʰě tɕʰug a ngě: pʰjǐn-gò]
2P stay.HON PART 1P-ERG go.PAST-MODAL
‘(Please) you stay, I will go.’ (Tournadre, 1991, p.101)

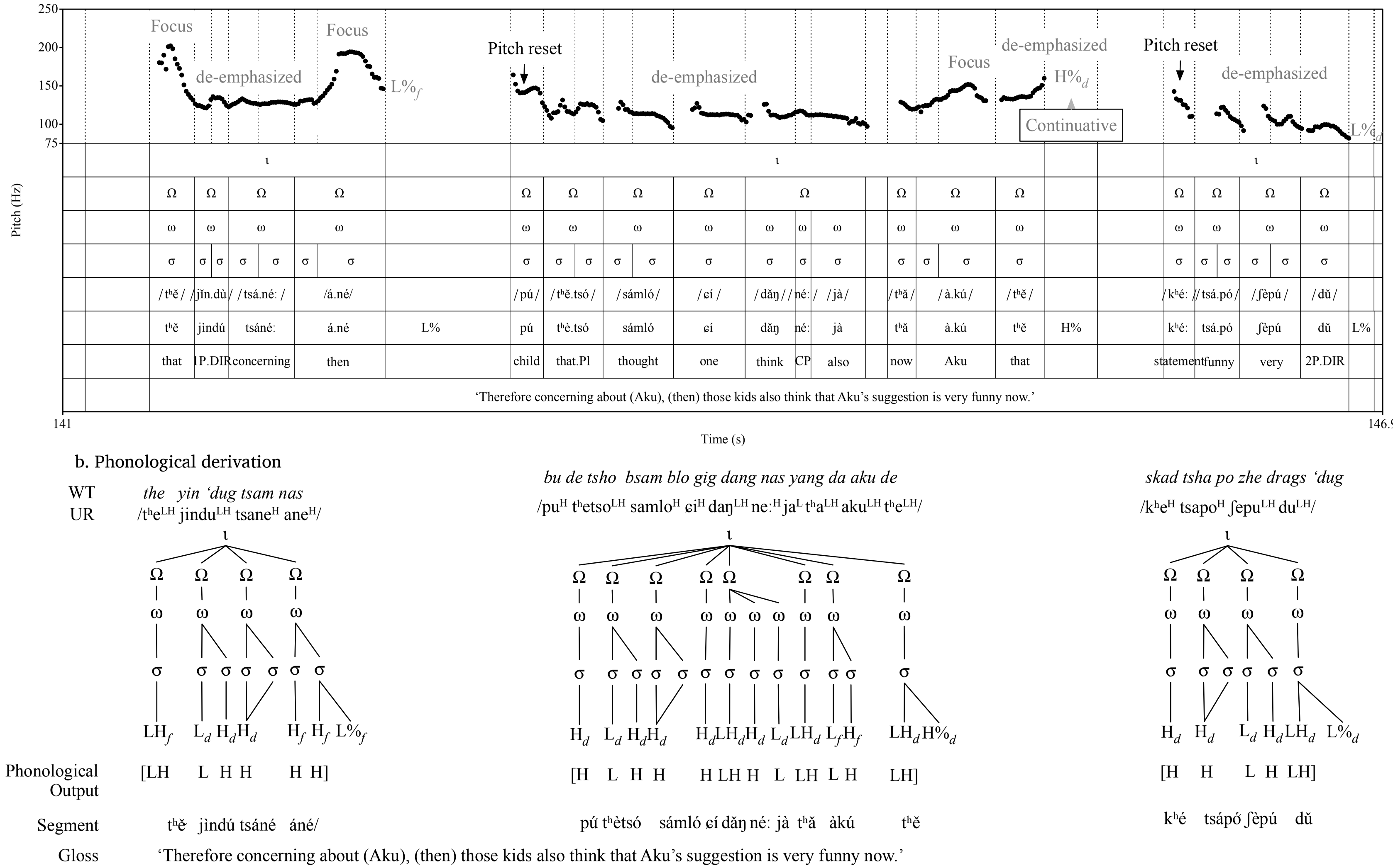


Figure 4.9: Example of post-focal de-emphasis.

Regardless of the sophisticated morphosyntactic devices used to encode information structure in LT, there is evidence that LT also utilizes prosody to apply focus to parts of an utterance. In LT, a focalized constituent tends to exhibit a raised pitch floor, wider and higher pitch range. This agrees with the *Effort Code* (Gussenhoven, 2004, p.85), where the speaker increases his or her articulation effort to optimize informational saliency. This typically results in greater amplitude, longer duration and, crucially, a raised and expanded pitch range. It seems that both lexical tones and boundary tones are affected when their TBU is under focus. I have formalized this with a subscripted "f" (e.g., L_f, LH_f, H_f, $L\%_f$, and $H\%_f$). Examples of tones under focused have already been seen in Figures 4.4 and 4.6.

Interestingly, focused constituents are often followed by a span of de-emphasis, or post-focal compression (PFC) (cf. Monich, 2011, p.39). Similar phenomenon is also observed in Amdo Tibetan (Wang et al., 2012), East Asian languages (Xu et al., 2004; Chen et al., 2009; Lee and Xu, 2010; Wang et al., 2011; Lee and Xu, 2012), and other languages (Xu and Xu, 2005; Zerbian et al., 2010; Xu, 2011; Taheri Ardali and Xu, 2012). Both the lexical and boundary tones of the de-emphasized prosodic constituent are marked with a subscripted "d" (e.g., L_d, LH_d, H_d, $L\%_d$, and $H\%_d$). The F0 of H and H% tones in the post-focal span of de-emphasis can be compressed to the point of reaching the level of L tones in some speakers, while L and L% tones are less affected as their F0 are already close to the baseline of the pitch range. The effect of post-focal de-emphasis can be clearly seen in Figure 4.9. For example, the pitch range of *tsam nas*/tsa^{H}ne^H/ 'concerning' in the first intonational phrase is severely compressed. Furthermore, the de-emphasis can be extended to the following intonational phrase as shown in the second and third intonational phrases in Figure 4.9.

Since subjects and topicalized constituents are often under focus, one can get the impression that there is phrase-initial raising in LT. Although this requires more work, I argue for the time-being that this is epiphenomenal and that normal focus (marked with subscript f) is sufficient. Note that phrase-initial focus often obscures the effect of pitch reset, as both phenomena tend to raise F0 in the same direction.

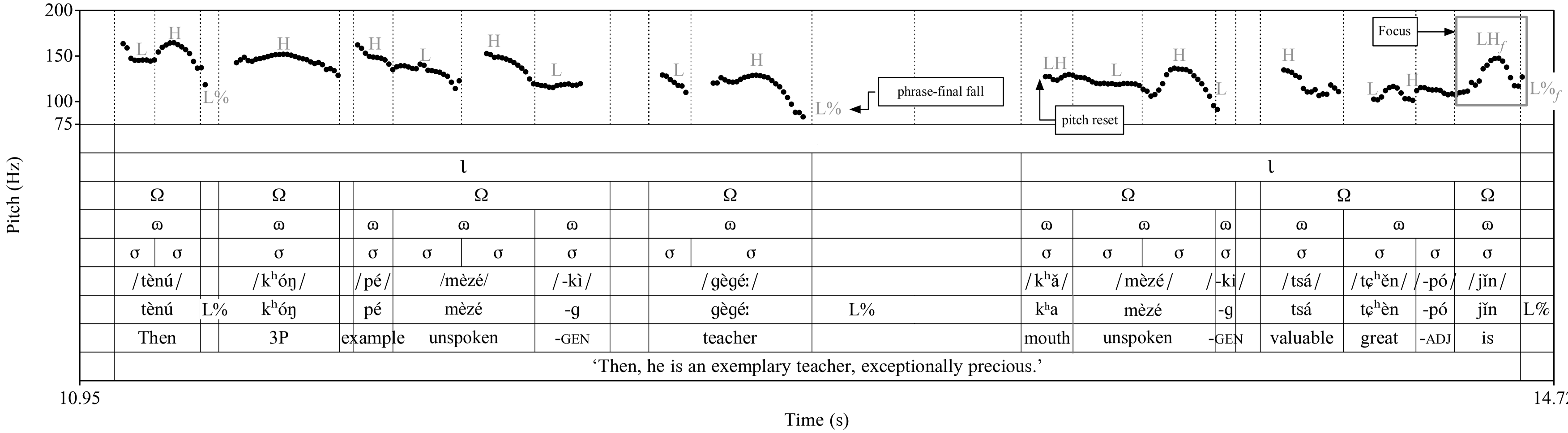
The LT information structure, thus, modulates the intonational contour of affected constituents. This is in contrast to lexical and boundary tones which attach themselves to a TBU (i.e., syllable) to form the overall pitch contour of the utterance. In what follows, I present two examples illustrating intonation being used for focus. The first example is given in (17 and 18) and Figure 4.10.

(17)	*de dus*	*khong*	*dpe*	*med zer ki*	*dge rgan*	[tènú k^{h}óŋ pé mè-zé-g gègé:]
	then	3P	example	NEG-speak-GEN	teacher,	

(18)	*kha*	*med zer ki*	*mtshan.chen.po*	yin	[k^{h}ǎ mè-zé-g tsá-tɕ^{h}è-pó jǐn]
	mouth	NEG-speak-GEN	valuable-great-ADJ	is	

‘Then, he IS an exemplary teacher, exceptionally precious.’

In (17 and 18) the phrase-initial discourse particle *de dus* /tènú/ ‘then’ cues the listener that new information is expected, but the phrase-final copula *yin* /jǐn/ is focused to emphasize that the preceding information is important. The effect is similar to the English: ‘Then, he **IS** an exemplary teacher, exceptionally precious.’ This utterance-final focus is strong enough to cancel out the effect of declination in 4.10.



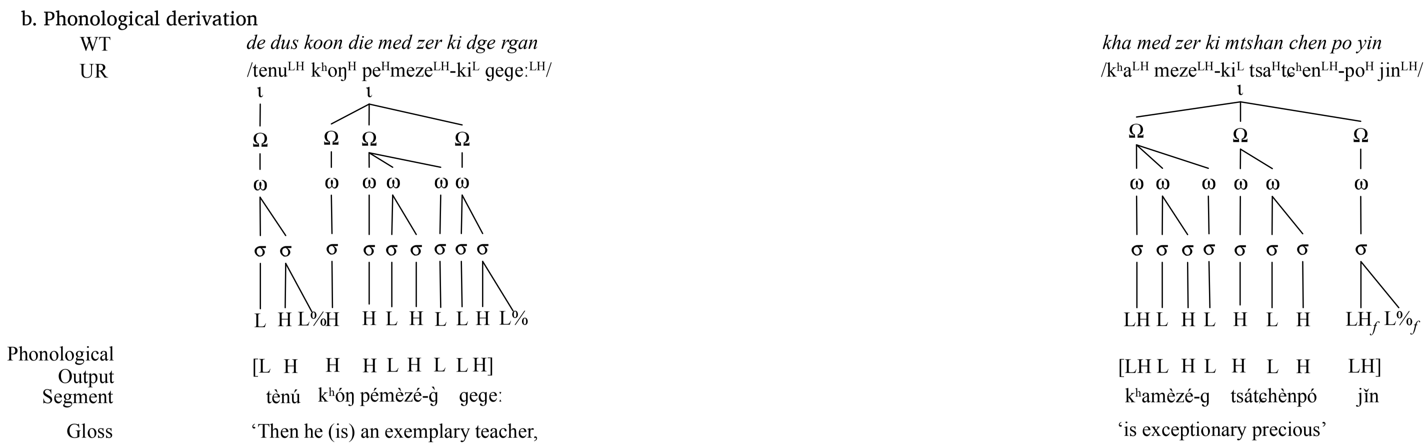


Figure 4.10: An example of a declarative statement extracted from a monologue.

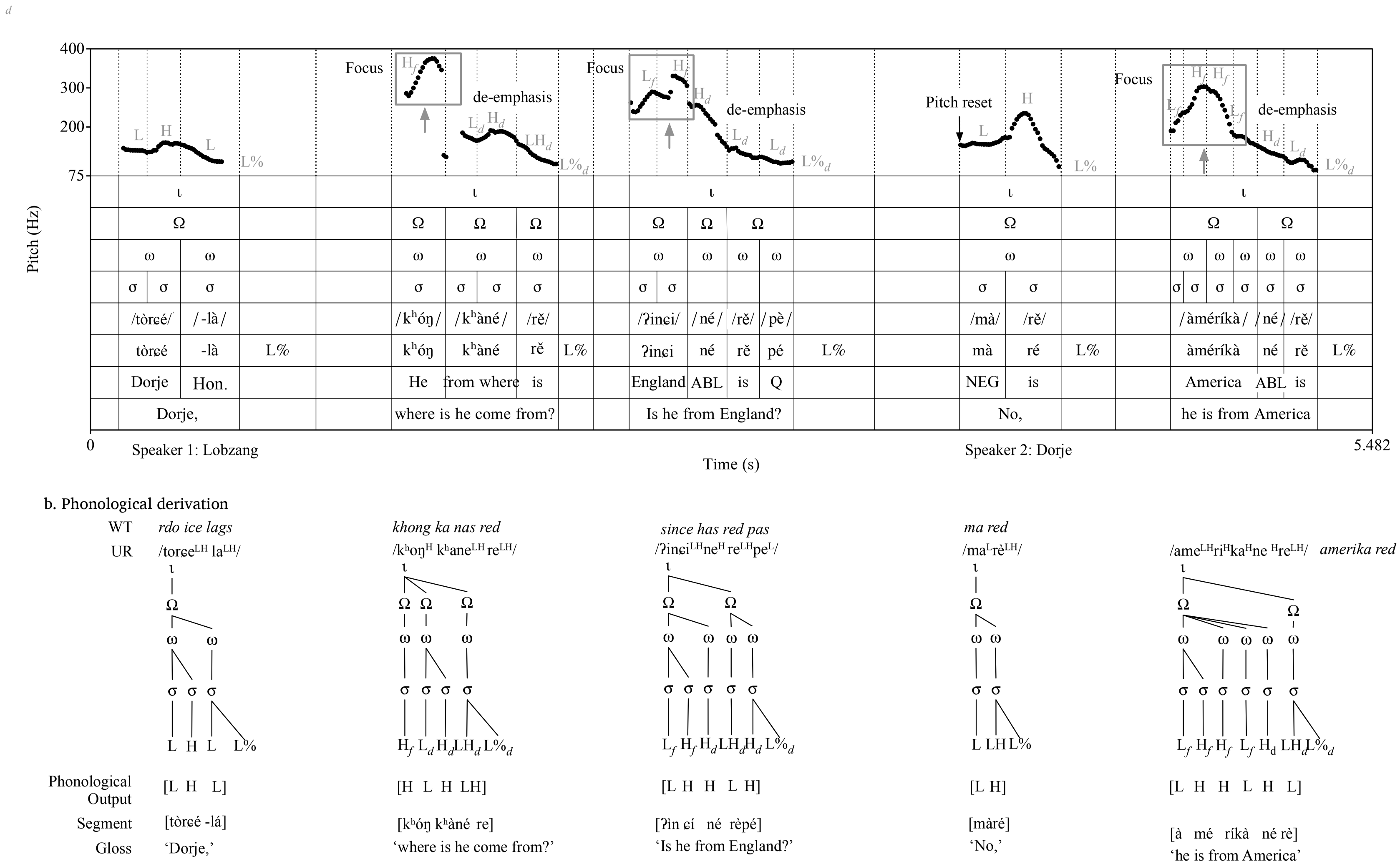


Figure 4.11: A dialogue between two speakers (Tournadre and Dorje, 2003, p.150).

A second example of focus is extracted from a dialogue between two speakers[4]:

(19) *rdo rce lags khong ka nas red* [tòrçé-la, kʰóŋ kʰàné-rè,]
Dorje-Hon., 3P from-where-is
'Dorje, where is he from?' (Tournadre and Dorje, 2003, p.150)

(20) *yinci nas red pas* [ʔìnçí-né rè-pè]
England-ABL is-Q
Is (he) from England' (Tournadre and Dorje, 2003, p.150)

(21) *ma red khong a.me.ri.ka nas red* [màré, kʰong àméríkà-néʔ rèʔ]
No, 3S America-ABL is
'No, (he) is from America' (Tournadre and Dorje, 2003, p.150)

The intonational pattern and phonological derivation of this sentence are presented in Figure 4.11. There are three focalized constituents in this dialogue fragment. All of these constituents happen to occupy the phrase-initial position, which also corresponds to the main topic. The first focus constituent is *khong* [kʰóŋ] '3P,' which is the topic of this dialogue. The H_f elevates the pitch floor of *khong* [kʰóŋ] '3P,' which also bears an H lexical tone. This effect creates the highest pitch peak in the intonation pattern of utterance (19 and 20) and marks it as the most prominent word. After that, the first speaker presents a Yes/No question, where the suggested country of origin, *yinci* [ʔìnçí] 'England,' is also contrastively focused as seen from its significantly high pitch onset and large pitch excursion. The second speaker then applies contrastive focus to *amerika* [àmérìká] 'America,' making use of the same acoustic cues. Post-focal compression (or de-emphasis) is also obvious in these examples: each focused constituent is followed by several syllables of compressed tonal targets (marked with a subscript d).

The examples presented here are not an exhaustive analysis of the prosodic ap-

[4]The recording of this dialogue is extracted from the CD that comes with Tournadre and Dorje (2003) manual of Standard Tibetan.

paratus employed to indicate the information structure in LT. Further research is warranted on the interaction between the morphosyntactic apparatus and the prosodic apparatus of information structure.

4.4 Spelling chant melody

In this section, I will provide a brief description and example of the spelling chant melody of LT, or 'sbyor klog' (Rose, 2009). This chant melody is usually acquired during the early stages of traditional Tibetan education, where every student is required to spell out a given word using this specific chant melody.

I encountered the spelling intonation while eliciting monosyllabic words from one of the participants (F1)[5]. In a speech sample consisting of a sequence of three tokens of the same word, the lexical tone of the first and third token is preserved, but the lexical tone of the second token is apparently replaced by an L tone. Also, the pitch floor of the first and third tokens are significantly elevated, as compared to the second token. Other participants also unconsciously applied this spelling intonation in few rare occasions, but not as pervasively as F1. An example of her intonational pattern is shown in Figure 4.12 and compared to a more typical elicitation intonation pattern produced by M1 in Figure 4.13[6].

[5]F1 tried to spell out each word with the spelling intonation at her first attempt. Interestingly, she uttered the same sequence of words with the identical spelling intonation after correction.

[6]Pauses between tokens have been shortened for illustration purposes.

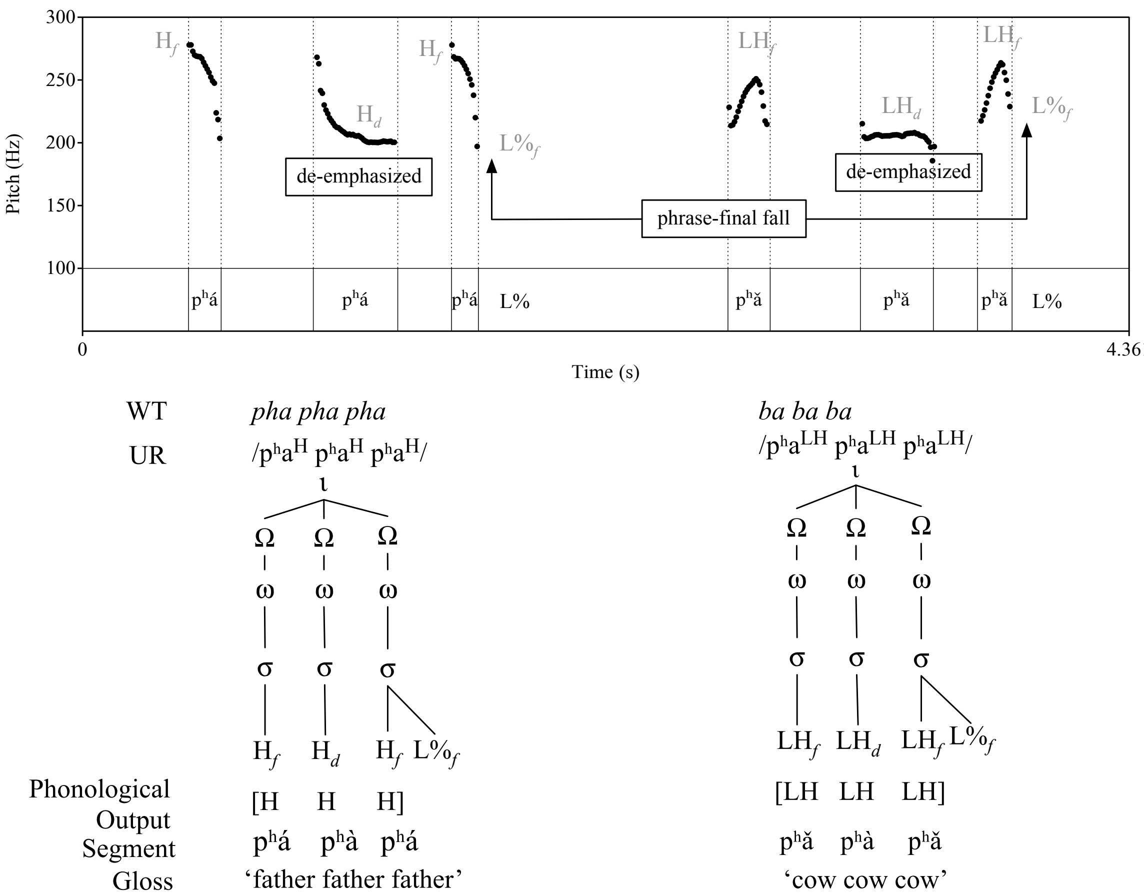


Figure 4.12: Examples of the spelling intonation of LT.

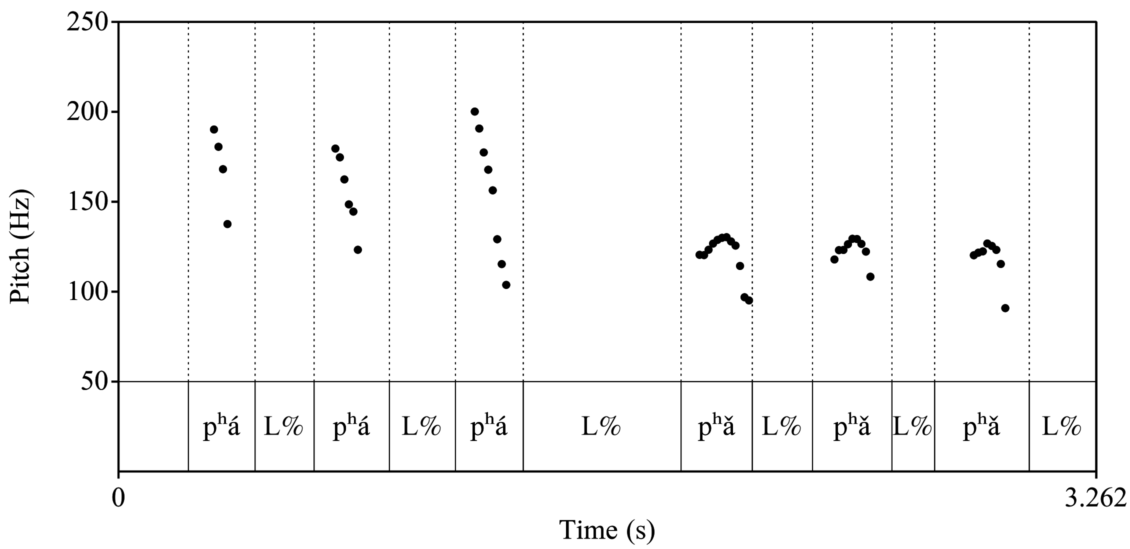


Figure 4.13: Example of normal elicitation intonation produced by M1.

I propose that the realization of the spelling intonation is achieved through the application of focus on a non-continuative statement that has a L% boundary tones. In other words, the phrase-initial and the phrase-final word are focused, which elevates the pitch floor and expands their pitch range (c.f. §4.3). The word tone in the second syllable, however, is de-emphasized and realized as with a low-level pitch. In addition, it also has a longer duration than the adjacent tokens.

I speculate that the expanded pitch range of focused words is to increase the saliency of the pitch contour for the language learner. In contrast, the segments of the second token are highlighted by de-emphasizing the word tone and increasing the duration. Thus, the spelling intonation provides enhanced contrast between word tones and segments, which facilitate the acquisition of LT orthography.

4.5 Tonal inventory of Lhasa Tibetan

This section presents both the lexical and intonational tonal inventory of LT, as shown in Table 4.2. The table provides a concise description of the surface realizations and transcriptions used in this dissertation. Also, the table includes the tonal alignment and its usage in the word tonology and intonational phonology of LT.

Surface Realization	Transcription	Alignment	Usage
Low tone	L	syllable	1. Syllable with the L of a decomposed LH 2. Affixes/Clitics with a lexical L
Rise	LH	syllable	LH of a monosyllabic prosodic word
High tone	H	syllable	1. Monosyllabic word with H 2. Syllable with the H of decomposed LH
High tone	!H	syllable	Downstepped H
Mid tone	M	syllable	After tone merger of a stem and a following particle whose segmental content is completely deleted /H + L/→M (rare, see §3.2.1)
Falling	L%	phrase-final syllable	*Phrase-final falling*
Rising	H%	phrase-final syllable	*Phrase-final rising*
Raised pitch floor and expanded pitch range	X_f	focus domain	Focused constituents
Compressed pitch range	X_d	post-focal domain	De-emphasized constituents

Table 4.2: Tone inventory of LT (“X” represents any LT tone.)

Chapter 5

F0 Synthesis Model of LT

The word tonology and intonational phonology of LT described in this thesis are based on the observation and analysis of the pitch pattern of words and utterances (cf. Chapter 3 and Chapter 4). In the framework of autosegmental phonology, the pitch contour is derived from the phonological output of an underlying representation. In LT, this can be either the lexical tones (e.g., L, LH, H) or the boundary tones (e.g., L%, H%). However, the surface phonological representation of an utterance needs to be phonetically implemented, a step that is often omitted in intonational models of specific languages. In the best cases, this results in analyses that are incomplete or difficult to evaluate (in some case, this is a convenient way to brush aside difficult issues).Thus, I propose to utilize an F0 synthesis model with explicit phonetic implementation rules to evaluate the adequacy of the proposed word tonology and intonational phonology of LT. The correlation between the synthesized F0 and the original F0 will then be compared. If the phonology is adequate, a high correlation is expected.

In the following sections, I outline the necessary criteria for an adequate F0 synthesis of LT (§5.1), give an overview of various existing F0 synthesis models (§5.2), and justify the selection of the F0 synthesis model best suited to handle the LT data (§5.3). Next, I provide a detailed description of the selected model: qTA (§5.5). Then, I present its application to LT (§5.5). Finally I describe the three main components of the model: the phonological component (§5.5.1), the F0 synthesis component (§5.5.2), and the evaluation component (§5.5.3).

5.1 Criteria for an adequate F0 synthesis model

The following are the design criteria for the F0 synthesis model of LT. First, the model should be able to represent the perceptually distinguishable pitch patterns of LT. Next, the model should automatically generate the pitch contour of an utterance, based on its phonological output. Also, the pitch contour synthesis component of the model should emulate the major physiological and physical constraints involved in pitch realization. For example, the model should factor in the greater difficulty in producing a rapidly rising F0 relative to a rapidly falling F0 and F0 target declination (Gussenhoven, 2004, p.98). Lastly, the model should be easy to implement in a commercial or a public-domain computational mathematical package such as MATLAB®.

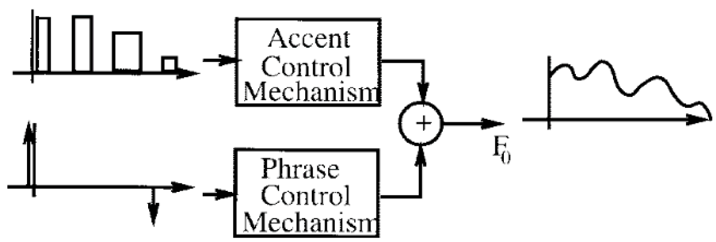
For practical reasons, two factors known to affect F0 contours will not be integrated in this synthesis model. First of all, the effect of emotions on intonation will not be encoded. The emotion of the speaker during utterance can have a significant effect on pitch floor and the duration of the syllable. For example, if a speaker is excited during an utterance, he tends to have larger pitch excursion and faster speech rate. While they could in theory be included in a model of F0, emotions are difficult to elicit in a natural way in controlled speech and hard to identify and classify in spontaneous speech. They are therefore beyond the scope of this model (and of the dissertation). Second, the effect of micro-prosody (e.g., intrinsic F0 of consonants and vowels) on pitch realization will not be integrated in the model. The reason for overlooking micro-prosodic effects is that it should be possible to test the ability of the phonological model developed in previous chapters to predict the general intonational contours of LT without encoding fine-grained micro-prosodic effects.

5.2 Overview of various F0 synthesis models

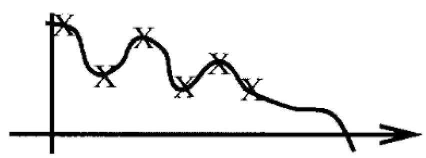
Currently, there are multiple F0 generating models described in the literature. The main difference between the various models is how two basic questions are answered: *how are F0 targets determined?*, and *how is F0 generated?*.

F0 targets can be obtained through either a phonological or a functional approach. In the former approach, surface phonological labels (word and boundary tones in LT) are converted into F0 targets. This approach is also known as the phonological F0 generation system (Sun, 2002). Currently, there are two prominent candidates in this category: the approach developed by the Institute of Perception (IPO) ('t Hart et al., 1990), and the Autosegmental Metrical (AM) phonology (Pierrehumbert, 1980; Ladd, 1996). The latter approach is adopted in this dissertation. In contrast, the functional approach model does not require prosodic labels to generate the intonation contour. These approaches obtain F0 targets by either modelling physiological aspects of the speech organs (Prom-on et al., 2009) or by using a parametric method. The F0 targets of the parametric method are obtained from statistical models trained on corpora, as in Hidden Markov Model (HMM) synthesis (Huang et al., 1990; Yu and Young, 2010; Lu et al., 2011, 2013). An HMM model would analyze a large corpus of data and try to construct an algorithm that can predict and automatically assign prosodic labels to the intonational pattern of an unknown utterance.

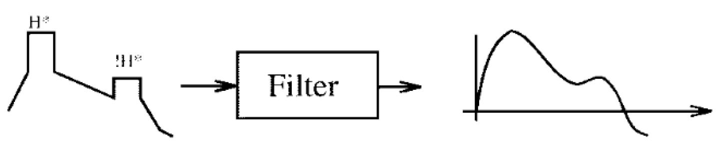
The next step consists in generating pitch contours based on the labels that are pre-specified or are inferred by the model. F0 generation is based on one of three techniques: a superposition of various pitch contours, an interpolation between F0 targets, or an hybrid of these two methods (Ross and Ostendorf, 1999). These techniques are illustrated in Figure 5.1.



(a) Superposition method



(b) Pitch target interpolation method



(c) Hybrid method

Figure 5.1: F0 generation techniques (Ross and Ostendorf, 1999, p.297.)

Examples of the superposition method (Figure 5.1a), are the command-response and the Fujisaki models (Öhman and Lindqvist-Gauffin, 1965; Fujisaki and Nagashima, 1969). In these models, the F0 contour is contributed from the following components: global pitch contours (e.g., intonational tones, phrase tone); local pitch contours (e.g., lexical tone, stress, or pitch accents); segmental effects (e.g., intrinsic F0 of segments); and information effects (e.g., focus, topic, giveness etc.) (Ross and Ostendorf, 1999; Prom-on et al., 2009). The amplitude of each effect is added to a baseline fundamental frequency, either as an enhancement (above F0 baseline) or a suppression (below F0 baseline). Subsequently, the integrated F0 contour is passed through a filter function to smooth out the F0 target values.

In the F0 target interpolation technique, as shown in Figure 5.1b, the model interpolates an F0 curve between a series of F0 targets indicated as 'X' in the figure. The interpolation can be a simple parabolic transition function, a spline function, or even a sophisticated prediction from a neural network model. The following are examples of the F0 synthesis model which utilize the F0 target interpolation technique: Spline (Hirst and Espesser, 1993; Hirst, 2005), Tilt/RFC (Taylor, 2000), Stem-ML (Kochanski and Shih, 2003), and TA/qTA/PENTA (Xu and Wang, 2001; Xu, 2005; Prom-on et al., 2009).

In the hybrid technique, as shown in (c) of Figure 5.1, a F0 target interpolation technique is applied to obtain a rough outline of the intonational contour, and this is then passed through a filter function to smooth the target values. The F0 targets of this rough outline are defined sequentially from a phonological analysis, like the Autosegmental Metrical Model. The filter function is similar to those used in the superposition model for F0 synthesis, hence, the hybrid approach. This hybrid technique can be found in the rule-based parametric F0 model proposed by Anderson et al. (1984) and the dynamical system model of F0 and Energy proposed by Ross and Ostendorf (1999).

5.3 Selection of the F0 synthesis model

In order to be able to justify the choice of a specific model for LT F0 resynthesis, I will introduce a few key notions by discussing the synthesis of English intonation proposed by Pierrehumbert (1981), which is closely associated with the Autosegmental Metrical (AM) model. As discussed earlier, the AM model, which is a phonological approach, assigns tone targets (lexical or boundary tones) to specific segmental anchors in an utterance. Pierrehumbert (1981, p.47-54) then proposes two tonal implementation rules for F0 generation. The first rule phonetically evaluates each tonal target and assigns it an F0 target based on the pitch range of the speaker. This approach is adopted in the F0 synthesis model of LT. The second rule realizes the pitch contour between two target values. There are two types of pitch contours: monotonic and 'sagging'. If one of the F0 targets is right on the minimum F0 of the speaker pitch range, a monotonic pitch contour is implemented. The monotonic pitch contour, thus, is either rising or falling. If both F0 targets are above the minimum F0 of the speaker's pitch range, the 'sagging' contour is implemented. The 'sagginess' of the F0 contour is determined based on the duration and the pitch difference between these two targets. The examples of computer-generated F0 contours in Figure 5.2 are reproduced from Pierrehumbert (1981). The left portion (I) illustrates the monotonic interpolation, and the right portion (II) illustrates the implementation of the 'sagging' contour. (II) also shows the difference in 'sagginess' between two more widely separated tone targets (a) as compared to the tone targets with a narrower gap (b). Pierrehumbert's second rule of AM F0 synthesis represents a functional approach that stylizes the F0 contour between F0 targets. This F0 synthesis technique is not ideal as it is not based on the physiologically constrained F0 synthesis model, which mimics the characteristics of the vocal cords during voicing. Consequently, it is necessary to identify an alternative pitch contour synthesis method if we are to use the framework of the Autosegmental Metrical model to build a realistic model of F0 implementation.

As far as I know, only three models of F0 synthesis mimic the physiology of the human vocal folds: Öhman and Lindqvist (Öhman and Lindqvist-Gauffin, 1965), Command Response/Fujisaki (Fujisaki and Nagashima, 1969), and qTA (Prom-on et al., 2009). The former two approaches use the superposition method for F0

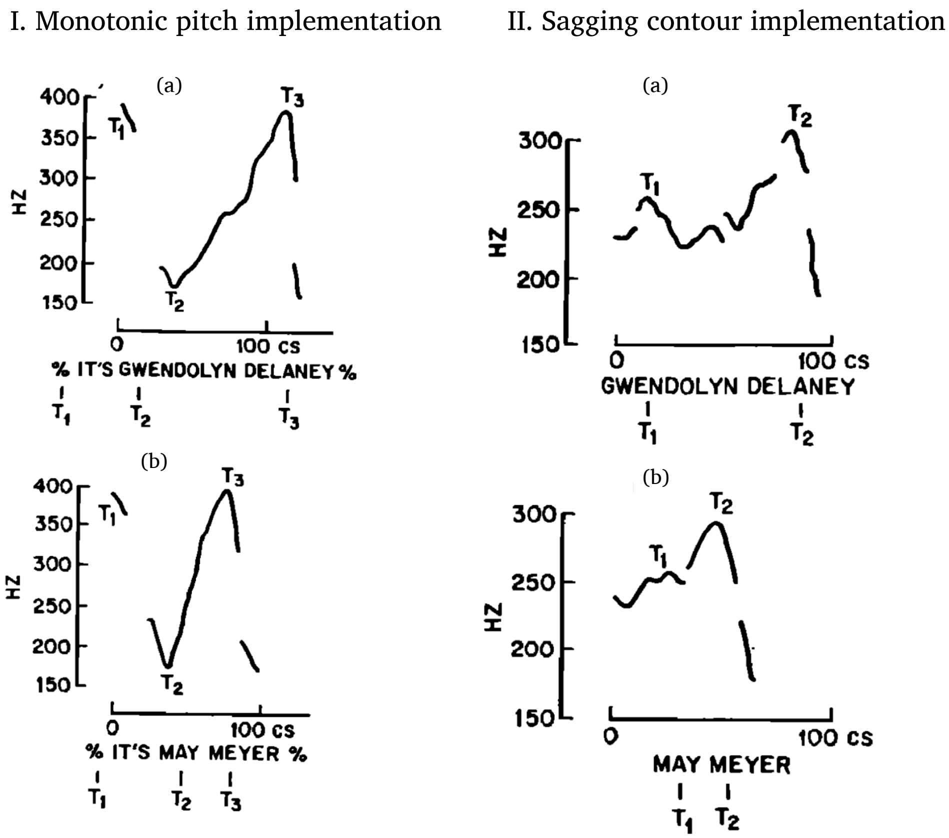


Figure 5.2: Tonal implementation rules of Pierrehumbert's AM Model (Pierrehumbert, 1981).

generation, which is the superposition for various levels (both global intonational and local tonal components) of pitch contours to produce an overall pitch contour. This approach is incompatible with at least two precepts of the AM framework: *the linearity of tonal structure*, and *the local sources for global trends* (Ladd, 1996, p. 42-43). The qTA approach does not have this difficulty as it is independent of the phonology. In the current dissertation, I have chosen to integrate the qTA pitch contour synthesis model with the autosegmental metrical theory to construct the model for LT (c.f. §5.5). I will provide a more in-depth discussion in the following section.

5.4 Quantitative Target Approximation (qTA)

The target approximation (TA) model and its quantitative model (qTA) serve as the fundamental framework for the F0 synthesis model of LT (Xu and Wang, 2001; Xu, 2005; Prom-on et al., 2009). To facilitate this discussion, I have reproduced a schematic illustration of the TA model in Figure 5.3.

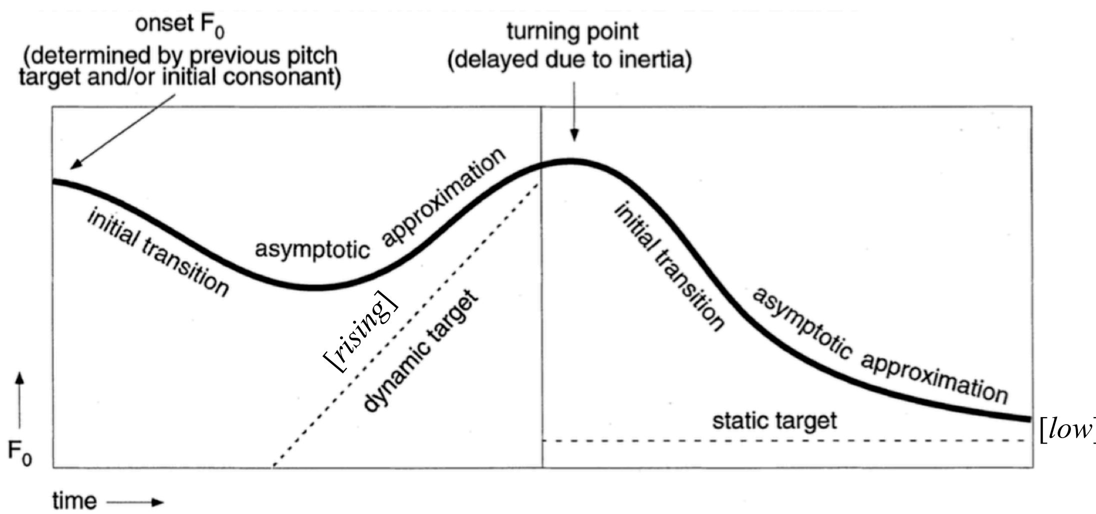


Figure 5.3: Schematic illustration of TA model

In this model, there are two main components: pitch targets and pitch contour generation. A pitch target is defined as "the smallest articulatorily operable units, such as tone and pitch accent" (Xu, 2005). There are two types of pitch targets: static and dynamic. A static pitch target can have a register specification, i.e., $[high]$, $[low]$, or $[mid]$. The $[low]$ pitch target is shown at the bottom of the right-hand panel of Figure 5.3. A dynamic F0 target is vectorial; it can be specified with either $[rising]$ or $[falling]$. A dynamic rising target is found in the left-hand panel of Figure 5.3.

A crucial aspect of the TA model is that targets are approximated rather than faithfully realized. The four rules define approximation are the following (Xu and

Wang, 2001, p.322) :

(22) a. A pitch target is implemented in synchrony with the host, i.e., starting at its onset and ending at its offset.
  b. Throughout the duration of the host, the approximation of the pitch target is continuous and asymptotic.
  c. A falling F0 movement is implemented faster than a rising movement.
  d. A pitch target containing a high pitch point is implemented with a higher F0 peak when followed by a pitch target containing a low point than when followed by a pitch target with no low point.

TA was later developed into the parallel encoding and target approximation (PENTA) model, which can incorporate other pitch-based communication meanings (Xu, 2005). The schematic of the model is reproduced in Figure 5.4, for the benefit of the reader:

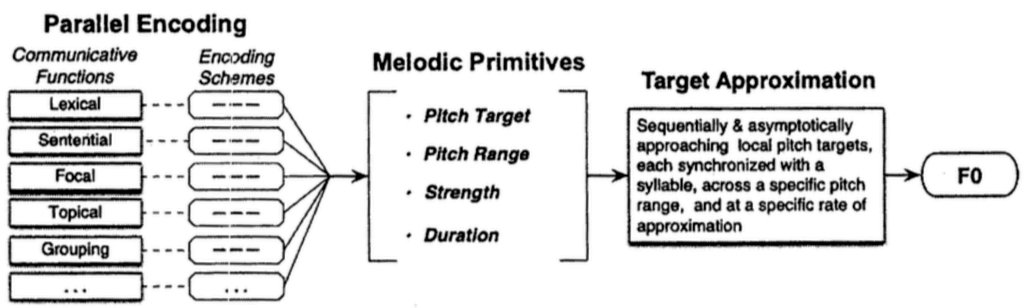


Figure 5.4: Schematic of PENTA model.

PENTA was then further developed into a quantitative version: (qTA). A block diagram representing qTA is also reproduced in Figure 5.5.

The F0 generation algorithm of qTA is a third-order critically damped linear system as shown in Table (5.1) (Prom-on et al., 2009, p.410). It emulates the characteristics of vocal fold vibrations during an utterance. This is also the component that I adopted for the F0 synthesis model of LT.

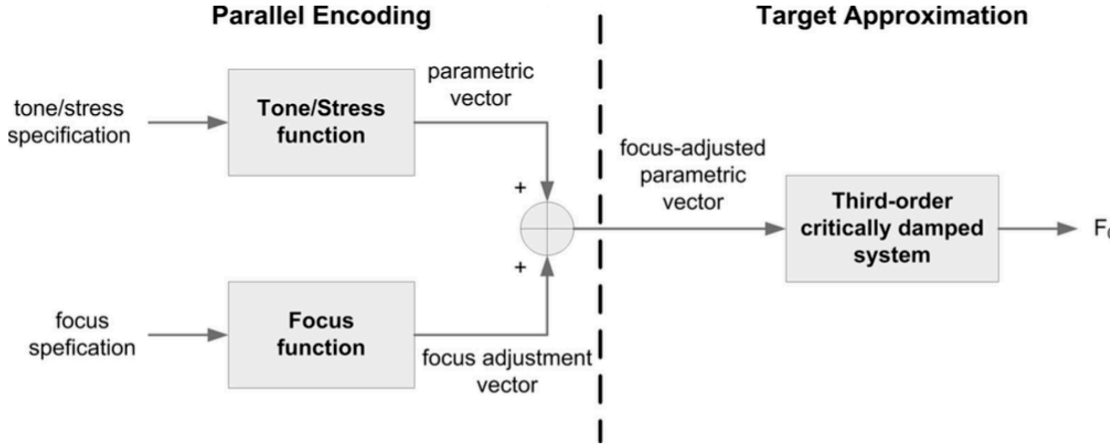


Figure 5.5: Block diagram of the qTA model (Prom-on et al., 2009, p.413).

$f_0(t) = x(t) + (c_1 + c_2t + c_3t^2)e^{-\lambda t}$	$x(t)$ = pitch target; λ = target approximation rate
a. $c_1 = f_0 - b$	f_0 = Initial pitch; b = static pitch target
b. $c_2 = f_0^{'} + c_1\lambda - m$	$f_0^{'}$ = Initial velocity; m = dynamic pitch target
c. $c_3 = (f_0^{''} + 2c_2\lambda - c_1\lambda^2)/2$	$f_0^{''}$ = Initial acceleration

Table 5.1: The algorithm of vocal fold vibration proposed by qTA.

This linear function, which represents the vocal vibrations of the vocal folds, returns to equilibrium as quickly as possible without oscillation. It represents the pitch target $x(t)$ and the transition of pitch contour from the current state towards the pitch target ($(c_1 + c_2t + c_3t^2)e^{-\lambda t}$). The transition of the pitch contour includes the parameter of the initial state that precedes the pitch target, which includes f_0 initial pitch, $f_0^{'}$ initial velocity, and $f_0^{''}$ initial acceleration.

The target approximation rate, λ, is the key constant that controls the damping characteristic of the vocal-fold vibration. It represents the articulatory effort required to reach a pitch target and must be speaker-specific to a certain extent (because of anatomical differences between speakers). The target approximation rate is one way to guarantee that, as in real life, speakers will not systematically achieve the planned targets because of co-articulatory constraints and speech rate.

I have adopted the qTA as the F0 synthesis module in the current dissertation because it is the most phonetically grounded method for F0 generation, because it is relatively easy to implement in the context of this research and because it allows future incorporation of other pitch-based communication functions, such as emotions and micro-prosody. However, some parameters of the qTA are modified. The implementation of qTA to LT is further discussed in §5.5.

5.5 LT F0 synthesis model (LTF0)

Based on the aforementioned design criteria (see §5.1), the LT F0 synthesis model (LTF0) consists of three main components: *a*) a tone target component, *b*) an interpolation component, and *c*) an evaluation component. These various components are illustrated in Figure 5.6. LTF0 is constructed in a *MATLAB_R2016a*® platform[1]. I will provide a detailed discussion of each component in the following sections.

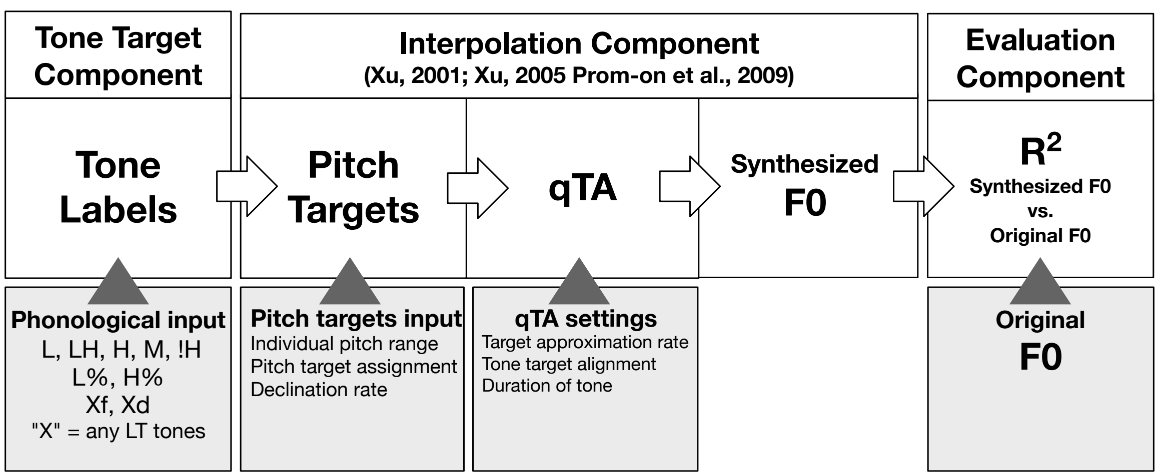


Figure 5.6: Components of LT F0 synthesis model (LTF0).

[1]*MATLAB_R2016a*® is a platform developed by MathWorks for computational mathematics with built-in graphic and programming capabilities. For more information, please visit the website: www.mathworks.com

5.5.1 Tone target component

The phonological input for the phonological component is extracted from manually annotated Praat textgrids containing surface tone targets and segments[2]. An example of the input is provided in Table 5.2[3].

Tone	WT	Syllable	Gloss
L	*de*	[te	'then'
H	*dus*	nu]$_\omega$	
H	*khong*	[kʰoŋ]$_\omega$	'he'
H	*dpe*	[pe]$_\omega$	'example'
L	*med*	[me	NEG
H	*zer*	ze]$_\omega$	'speak'
L	*ki*	[gi]$_\omega$	GEN
L	*dge*	[ge	'teacher'
H	*rgan*	gen]$_\omega$	
L%			
LH	*kha*	[kʰa]$_\omega$	'mouth'
L	*med*	[me	NEG
H	*zer*	ze]$_\omega$	'speak'
L	*ki*	[-ki]$_\omega$	GEN
H	*mtshan*	[tsa]$_\omega$	'valuable'
L	*chen*	[tɕen	'great'
H	*po*	po]$_\omega$	ADJ
LH_f	*yin*	[jin]$_\omega$	'is'
L_f%			

Table 5.2: An example of phonological input for the F0LT model

The tonal labels that can be assigned to individual syllables are shown in Table 5.3. They are all surface tone targets described in previous chapters. They include the four surface tones (L, LH, H and M), the downstepped tone (!H), the surface tones on focused syllables (L_f, H_f, LH_f), and the tone of the de-emphasized syllable, which has a subscript "d" (L_d, H_d, LH_d, etc.). This inventory also includes boundary tones (L%, H%).

[2]The MATLAB can extract all the data from all the relevant files in a directory in one go.
[3]Square brackets, '[' or ']', represent the boundaries of a prosodic word.

Tone label	Description
Surface tones	
L	low level tone
LH	LH rising contour
H	high level tone
M	mid level tone
!H	downstepped high tone
Boundary tones	
L%	phrase-final L boundary tone
H%	phrase-final H boundary tone
Focused tones	
X_f	focused tone
De-emphasized tones	
X_d	de-emphasized tone

Table 5.3: Recognized tonal labels in F0 synthesis model of LT ("X" represents any LT tones.)

5.5.2 Interpolation component

The interpolation component corresponds to the qTA algorithm. In this section, I will discuss the required inputs and settings for this component. If the goal of the synthesis is to compare the intonation of a real speaker with a simulation of his/her F0, it requires five inputs. Two are speaker specific: the F0 range and the declination rate. The other three are fixed for all speakers: F0 targets corresponding to tonal targets, target approximation rates (λ), and F0 target alignment values (FTA). Otherwise, based on generic settings and inputs, the F0 synthesis can produce a generic LT speaker. In the next sections, I will show how these parameters are set and how they can be largely derived from one another to avoid a proliferation of arbitrary decisions. I will then illustrate the LT F0 synthesis with real examples.

5.5.2.1 Individual pitch range

Every speaker has his or her own unique pitch range. The LT F0 synthesis model (LTF0) predicts the F0 targets assigned to each tonal target based on a speaker's individual pitch range. As we will see in the next few pages, the F0 of the tonal labels of an individual speaker is derived from the descriptive statistics of F0: the average F0 ($F0_{avg}$) and the standard deviation of the F0 ($F0_{sd}$). These values are based on a set of extracted pitch contours from speech samples (about 240 syllables, 3 tokens for each syllable, and 10 sampling points for each syllable) recorded from each LT participant. The pitch range indicators of the five speakers whose data was used for calibration of the LT F0 synthesis model are reported in Table 5.4. Note that speaker F1 is not included; as she is the only woman in the sample, it was impossible to determine if peculiarities in her data are characteristics of female speech or idiosyncrasies. Therefore, the model developed here is only valid for male voices.

Participants	$F0_{avg}$ (Hz)	$F0_{sd}$ (Hz)	Declination Rate
M1	145	33	20 Hz/s
M2	112	21	10 Hz/s
M3	115	12	5 Hz/s
M4	151	23	20 Hz/s
M5	123	24	50 Hz/s

Table 5.4: Individual parameters for F0 and declination rate

5.5.2.2 Declination rate

Declination is the gradual F0 reduction during an utterance caused by decreasing sub-glottal pressure. The declination rate is speaker-dependent. There is a wide range of variation among LT participants, consistent with studies of non-Tibetan languages (Gussenhoven, 1984; Pierrehumbert and Beckman, 1988; Shih, 2000). I also observed a difference in declination rate between the H and L. It appears that H is more susceptible to the effect of declination as compared to L. The declination rate of H tones for a given LT participant is obtained empirically from the recorded corpus and shown in Table 5.4. The declination rate (Acc_declination) of the L tones is set to be half that of the H, which seems to roughly correspond to observed values in the corpus. For example, the declination of M1 is estimated to be around 20Hz/s for the Hs and about 10Hz/s for the Ls, starting from the onset of the utterance. In the current implementation, the declination rate (Acc_declination) is formulated as an accumulating lowering of the F0 target of a given syllable. The amount of F0 reduction depends on the position of the syllable in an utterance. Thus, I am expecting a significantly lower pitch floor at the end of a longer sentence.

5.5.2.3 F0 target assignment

According to the qTA model, the F0 target is defined as the "underlying goal" of the word tone or intonational tone (Prom-on et al., 2009, p.409). It is important to note that each tonal label has a predefined F0 target in LTF0, which is in contrast to the dynamic nature of the qTA model. This predefined F0 target reflects the categorical nature of this phonological output. A F0 target can be denoted by the simple linear equation as shown in (5.1).

$$x(t) = mt + b \tag{5.1}$$

As the implementation of the F0 target should occur on its tone-bearing syllable, time (t), corresponds to the beginning of the syllable. The m and b represent the slope and the apex of the F0 target, respectively called the static and dynamic F0 targets in the qTA model. A contention of my model is that only static targets are necessary in LT and that dynamic targets are superfluous. Thus, all the surface level tones ($L, M, H, !H$) and boundary tones ($L\%, H\%$) in my model have a static F0 target, which is defined by their height (b) alone. The contour tone LH is treated as the concatenation of L and H, and has two F0 targets: the L is the F0 target at the initial position, and the H is the F0 target at the offset of the pitch contour.

As mentioned earlier, individual F0 targets are derived from the pitch range and declination of a given LT speaker. The relative height of F0 targets is the same for all speakers in this model. The formulas for calculating the F0 targets of an individual speaker are given in Table 5.5 (Individual $\mathrm{F0}_{avg}$ and $\mathrm{F0}_{sd}$ and Acc_declination are given in Table 5.4)

Tone Label	F0 target	Algorithm
Surface Tones		
H	$\mathrm{F0}_{H}$	$= \mathrm{F0}_{avg} + 1 \times \mathrm{F0}_{sd} - \text{Acc\_declination}$
L	$\mathrm{F0}_{L}$	$= \mathrm{F0}_{avg} - 0.5 \times \mathrm{F0}_{sd} - 0.5 \times \text{Acc\_declination}$
M	$\mathrm{F0}_{M}$	$= \mathrm{F0}_{avg} + 0.35 \times \mathrm{F0}_{sd} - \text{Acc\_declination}$
!H	$\mathrm{F0}_{!H}$	$= \mathrm{F0}_{avg} + 0.35 \times \mathrm{F0}_{sd} - \text{Acc\_declination}$
Boundary Tones		
H%	$\mathrm{F0}_{H\%}$	$= \mathrm{F0}_{avg} + 1.5 \times \mathrm{F0}_{sd} - \text{Acc\_declination}$
L%	$\mathrm{F0}_{H\%}$	$= \mathrm{F0}_{avg} - 0.7 \times \mathrm{F0}_{sd} - 0.5 \times \text{Acc\_declination}$
Focused Tones		
H_f	$\mathrm{F0}_{Hf}$	$= \mathrm{F0}_{avg} + 2 \times \mathrm{F0}_{sd} - \text{Acc\_declination}$
L_f	$\mathrm{F0}_{Lf}$	$= \mathrm{F0}_{avg} + 0.5 \times \mathrm{F0}_{sd} - 0.5 \times \text{Acc\_declination}$
M_f	$\mathrm{F0}_{Mf}$	$= \mathrm{F0}_{avg} + 1 \times \mathrm{F0}_{sd} - \text{Acc\_declination}$
$!\mathrm{H}_f$	$\mathrm{F0}_{!Hf}$	$= \mathrm{F0}_{avg} + 1 \times \mathrm{F0}_{sd} - \text{Acc\_declination}$
$\mathrm{H\%}_f$	$\mathrm{F0}_{H\%}$	$= \mathrm{F0}_{avg} + 2.25 \times \mathrm{F0}_{sd} - \text{Acc\_declination}$
$\mathrm{L\%}_f$	$\mathrm{F0}_{H\%}$	$= \mathrm{F0}_{avg} + 0.3 \times \mathrm{F0}_{sd} - 0.5 \times \text{Acc\_declination}$
De-emphasized Tones		
H_d	$\mathrm{F0}_{Hf}$	$= \mathrm{F0}_{avg} - 0.5 \times \mathrm{F0}_{sd} - 0.5 \times \text{Acc\_declination}$
L_d	$\mathrm{F0}_{Ld}$	$= \mathrm{F0}_{avg} - 0.6 \times \mathrm{F0}_{sd} - 0.5 \times \text{Acc\_declination}$
M_d	$\mathrm{F0}_{Md}$	$= \mathrm{F0}_{avg} - 0.5 \times \mathrm{F0}_{sd} - \text{Acc\_declination}$
$!\mathrm{H}_d$	$\mathrm{F0}_{Hf}$	$= \mathrm{F0}_{avg} - 0.5 \times \mathrm{F0}_{sd} - 0.5 \times \text{Acc\_declination}$
$\mathrm{H\%}_d$	$\mathrm{F0}_{H\%d}$	$= \mathrm{F0}_{avg} - 0.4 \times \mathrm{F0}_{sd} - \text{Acc\_declination}$
$\mathrm{L\%}_d$	$\mathrm{F0}_{H\%d}$	$= \mathrm{F0}_{avg} - 1 \times \mathrm{F0}_{sd} - 0.5 \times \text{Acc\_declination}$

Table 5.5: Formulas for deriving individual F0 targets.

Here are plain word descriptions of the F0 targets for a few tonal targets. The

F0 target of H is 0.5 standard deviations above the average F0 of the speaker. Similarly, the F0 target of L is 0.5 standard deviations below the average F0 of the speaker. The F0 target of the boundary tones is at the extremes of the pitch range, where H% is 1.5 standard deviations above the average F0, and L% is 0.7 standard deviations below the average F0. The F0 target of the focused tones is significantly higher than their non-focused counterparts. The H_f is 2.0, and the L_f is 0.35 standard deviations from the average F0. The latter is about the same height as a downstepped H_f. Lastly, the F0 target of a de-emphasized H (H_d) is reduced to the same level as the L, but the F0 target of a de-emphasized L (L_d) is only slightly reduced to 0.6 below the average F0.

5.5.2.4 The setting of target approximation rate (λ)

The target approximation rate is the rate of F0 change that determines how fast a target can be reached. Thus, it is possible that the F0 target cannot be reached if the duration is too short or the approximation rate is too low. The target approximation rate in the current model is obtained via parametric testing. In general, the approximation rate of the level tone is greater than that of the falling tone, which is in turn greater than that of the rising tone (Xu and Wang, 2001). In this model, target approximation rates are determined based the difference between the F0 target and the initial pitch for each syllable of any given word. Their values, which are not speaker-specific, are given in Table 5.6. In plain words, the target approximation rates in Table 5.6 encode the fact that it is more difficult to raise F0 than to lower it, and that F0 rising/falling contours are more difficult to produce than F0 plateaus. Note that although I treat target approximation rates as fixed for all speakers in my basic model, it is likely that these values are actually speaker-specific due to the idiosyncrasies of laryngeal anatomy. I will briefly show how these values could be fine-tuned to better approximate the F0 contours of specific speakers at the end of this chapter.

Pitch contour	F0 target - Initial pitch	λ
Rising	>0	40 Hz/s
Falling	<0	60 Hz/s
Level	$=0$	80 Hz/s

Table 5.6: The algorithm for the determination of target approximation rate λ.

5.5.2.5 The setting of F0 target alignment (FTA)

The F0 target alignment (FTA) is another important constant in the F0 synthesis of LT. It is defined as the ratio between the time position of F0 target (T_{FTA}), (or the apex of the F0 contour from the beginning of the rhyme (t_0)), and the overall duration of the rhyme (t), as shown in the following equation (5.2).

$$FTA = \frac{t_{F0_{max}} - t_0}{t} \tag{5.2}$$

The duration of the rhyme (t) is one of the deciding factors in determining the F0 target alignment of F0 target. FTA determines the duration of a synthesized level tone that will be sustained within a rhyme in this model. Consequently, FTA also indicates the starting point of the synthesized F0 contour that is submitted to the effect of tonal coarticulation with the following tone in LTF0. The FTA value plays a critical role in F0 realization, but it is difficult to predict due to the influence and interaction of multiple factors. Such factors include the type of tone (L, H, LH), the segmental context, the phonological context, and the idiosyncrasies of a given speaker. Also, the FTA value will interact with other functions of this model – mainly the F0 target approximation rate (λ). Thus, it is difficult to have a fixed FTA for all different length of rhyme duration.

In (5.3), I proposed a simple function, $FTA(t)$, which provides different FTA

values based on rhyme duration (t). This FTA function reflects the fact that a LT speaker tries to reach the F0 target over a larger proportion of the rhyme when it is short and that there is more time for tonal co-articulation when the rhyme duration is longer. The particular slope of the function was determined by trial and error, using the evaluation component described in §5.5.3. The function plot of the equation is shown in Figure 5.7.

$$FTA(t) = (0.7 - \frac{t}{2}) \tag{5.3}$$

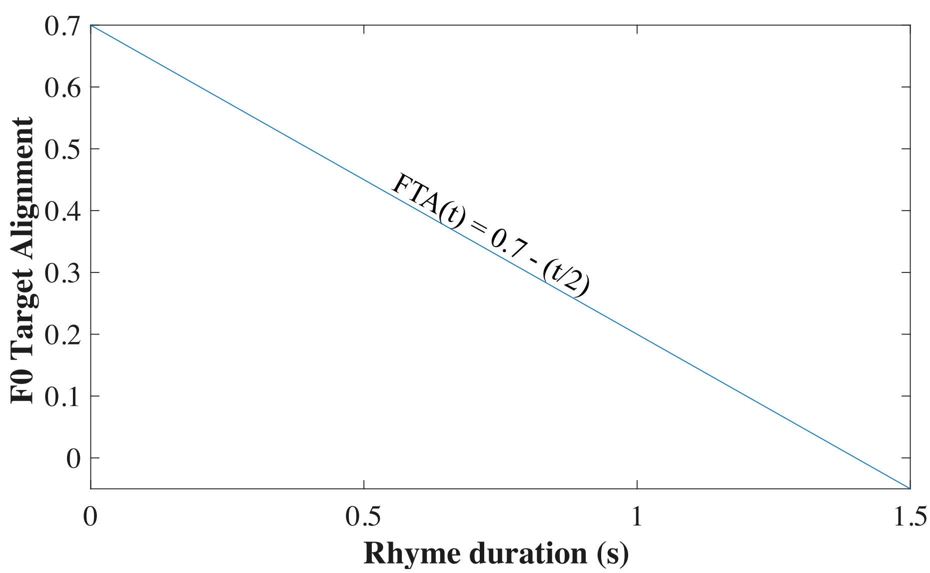


Figure 5.7: The function plot of FTA.

5.5.2.6 Examples of F0 synthesis

In the LTF0, F0 is synthesized one syllable at a time until every syllable of an utterance is processed. The rhyme of each syllable is divided into two portions separated by the F0 target alignment value (FTA). The first portion represents the pitch realization of the current tone (T_n) of the current syllable (σ_n), and the second portion represents the pitch realization of the tonal coarticulation between the current tone (T_n) of the current syllable and the next tone (T_{n+1}) of the following syllable (σ_{n+1}). Since the goal of the current model is to compare real speech with an F0 resynthesis, duration in the F0 synthesis is based on the duration of the original syllables produced by a given speaker. The duration of a given syllable is extracted from the original sound file via a text file generated by PRAAT. This text file provides the initial and final time position of each syllable in an utterance. Schematic representation of various tone targets are illustrated in Figure 5.8. These resynthesized F0 contours have a 0.5 FTA value for illustrative purposes; in real synthesized stimuli, the FTA would depend on rhyme duration.

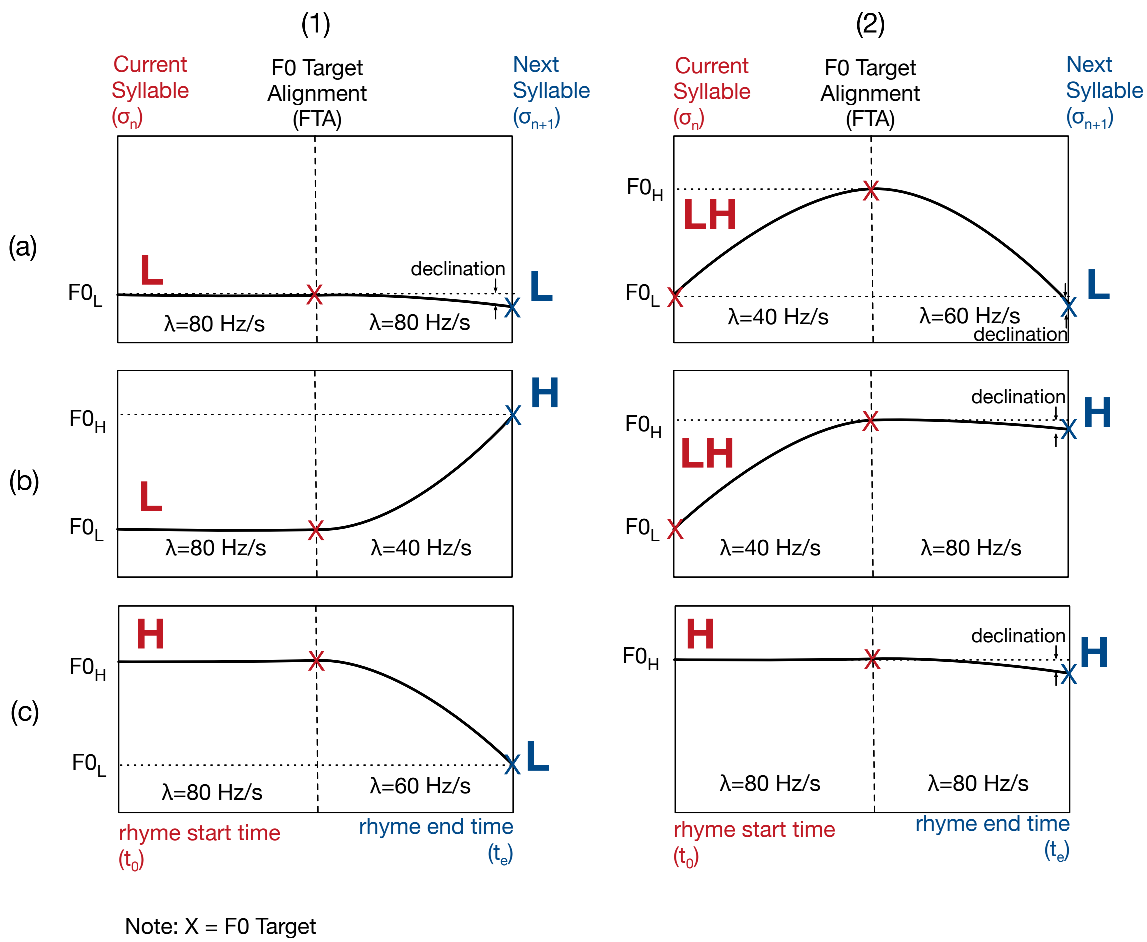
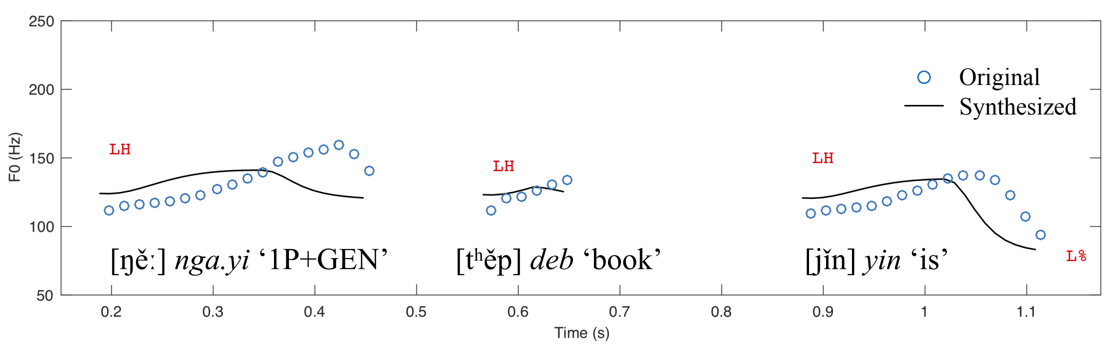


Figure 5.8: Schematic representation of the resynthesized F0 of various tonal targets.

In this qTA algorithm, each portion of the rhyme requires the input of an initial pitch, a F0 target, the F0 target of the following syllable, a target approximation rate, and a duration. The F0 contour is then obtained by using the qTA algorithm. The value of the initial pitch of the current tone portion that was used in this model is usually the same as the F0 target of a particular tone label with the exception of LH. For example, the initial pitch of an H tone label has an H pitch value ($F0_H$), which is the same as the H F0 target shown in (c2) of Figure 5.8. The initial pitch of an LH tone label is an L pitch value, ($F0_L$), followed by an H pitch value ($F0_H$), as shown in (a2) and (b2) of Figure 5.8. The initial pitch of the coarticulation portion is the F0 of the offset ($F0_{offset}$) of the previous portion, and the F0 target is defined by the tone label of the following syllable. In addition, the F0 value of the F0 target of the following syllable is decreased due to declination.

An example of the synthesized F0 contour of an entire utterance is given in Figure 5.9. We see that despite some discrepancies, there is a fair match between the natural and synthesized F0 contours.



Free Translation: “(This) is my book.”

Figure 5.9: Example of F0 synthesis of LT, speaker M1

5.5.3 Evaluation component

The evaluation function determines the adequacy of the model. This evaluation function plots both the F0 extracted from the model and the original human utterance. A correlation between the F0 of each glottal cycle in the original recording and the resynthesized F0 value at the corresponding time is then obtained. This can be used as a tool for fine-tuning the model, but is also as a way of assessing the explanatory power of the final model.

Participants	N	$F0_{avg}$	$F0_{sd}$	λ_{rising}	$\lambda_{falling}$	λ_{level}	R^2
M1	181	145	33	40	60	80	0.862
M2	138	112	21	40	60	80	0.657
M3	138	124	12	40	60	80	0.468
M4	128	151	23	40	60	80	0.270
M5	118	123	24	40	60	80	0.609

Table 5.7: The correlation and descriptive statistics of F0 for LT participants (Basic parameter.)

Participants	N	$F0_{avg}$	$F0_{sd}$	λ_{rising}	$\lambda_{falling}$	λ_{level}	R^2
M1	181	145	33	40	60	80	0.862
M2	138	112	21	40	60	80	0.657
M3	138	124	12	20	120	120	0.480
M4	128	151	23	40	60	80	0.289
M5	118	123	24	40	60	80	0.609

Table 5.8: The correlation and descriptive statistics of F0 for LT participants (M3's λs are adjusted.)

Participants	N	$F0_{avg}$	$F0_{sd}$	λ_{rising}	$\lambda_{falling}$	λ_{level}	R^2
M1	181	145	33	40	60	80	0.891
M2	138	112	21	40	60	80	0.656
M3	138	124	12	20	120	120	0.480
M4	128	151	47	40	60	80	0.403
M5	118	123	24	40	60	80	0.609

Table 5.9: The correlation and descriptive statistics of F0 for LT participants (M4's standard deviation of F0 is adjusted)

The R-squared values of various LT participants are shown in Table 5.7. The R-squared values in the table suggest that the model has a relatively good correlation for most participants (i.e., M1, M2, and M5), but not all. It should be emphasized that we are not aiming for a R-squared of 1 as the F0 synthesis model developed here does not try to include the effects of emotions and of microprosody. Yet, we should be able to obtain better R-squared for speakers M3 and M4. The R-squared value of M3 can be significantly improved by tweaking the λ as shown in Table 5.8. Recall from §5.5 that approximation rates should be speaker-specific to some extent, as speakers have different laryngeal anatomies, but that I have chosen to build the basic F0 resynthesis model with fixed values for the sake of economy. In practice, this comes down to saying that speaker M3 can drop his F0 faster than other speakers, but has a much harder time raising it. The R-squared values for M4, however, can only improved by increasing its $F0_{sd}$ (from 23Hz to 47Hz) as shown in Table 5.9. This probably means that the word list elicitation data upon which the F0 standard deviation of M4 was calculated is more monotonous, and therefore exhibits much less variation, than his spontaneous recordings. This suggests that in future work, F0 means and standard deviations should be calculated from spontaneous speech rather than formal word lists. It therefore appears that the correlation between synthesized and natural speech can be easily improved by modifying parameters that are known to be speaker-specific.

The R-squared values obtained in Table 5.9 are now sufficiently high to argue that the phonological model of tone and intonation upon which the F0 resynthesis is based is representing the crucial elements of LT prosody fairly accurately. Three further developments are likely to allow me to raise the accuracy of the predicted F0 contours to a point where very natural F0 contours would be obtained (R-squared values similar to that of M1): the first one is the integration of micro-prosody, i.e., the small but pervasive F0 perturbations induced by different types of consonants and vowels. While it could take a considerable amount of time, it is just a matter of hard work. The other areas that need to be improved are the integration of emotions and of a more gradient encoding of focus and information structure. They are far less trivial as they will require the labelling and encoding of pragmatic and attitudinal information in the training datasets.

The integration of Autosegmental Metrical phonology and quantitative Target Approximation (qTA) for F0 synthesis has proven to be productive in the current dissertation. It substantiates the idea that a phonologically-based model can predict and synthesize the F0 of a native speaker with a relatively small number of pre-set parameters. Furthermore, the qTA pitch contour synthesis emulates the physiological characteristics of the human vocal cords, which is phonetically more realistic than an arbitrary interpolation between F0 targets. In essence, this integrated approach provides a more linguistically grounded alternative to the current mainstream, non-phonological or functional approaches to pitch contour synthesis.

Chapter 6

Conclusion

This dissertation provides a detailed description of the tonal and intonational phonology of Lhasa Tibetan (LT), based on data elicited and recorded from members of the Tibetan-Canadian community in Ottawa and Toronto. The first two chapters of the dissertation contain background information about Lhasa Tibetan, a summary of previous research on its tones and intonation and the theoretical framework and conceptual tools used in the rest of the dissertation (see §1 & §2). In this final chapter, I recapitulate the most important findings and claims made in other chapters. In §6.1, I review my take on LT word tonology and the claim that LT has an active stress system. In §6.2, I discuss the LT communicative functions and their interactions with information structure. In §6.3, I go over the main features of my LT F0 re-synthesis model, which supports the validity of the phonological analysis of LT word tonology and intonational phonology. Finally, §6.4 will present my plan for further research.

6.1 LT Word tonology

In Chapter 2, I establish that the prosodic structure of LT brings evidence for four main constituents at or below the word level: a) the mora encodes vowel length contrasts, b) the syllable is the tone-bearing unit (TBU), c) the prosodic word, which is non-recursive and maximally binary, delimits the application of most tonal processes, and d) the prosodic word group, which matches grammatical words, is the domain of downstep. The latter constituent is similar to what was recently established by other researchers for other languages (cf. Vigário, 2003, 2010). This prosodic structure provides evidence against the universality of the Prosodic Hierarchy (Selkirk, 2002; Nespor and Vogel, 2007) in that it has no phonological phrase, but has two word-level constituents. A polysyllabic word is thus considered to be a prosodic word group comprised of smaller prosodic words. The surface realization of polysyllabic, trisyllabic, and quadrisyllabic words is a concatenation of surface realizations of monosyllabic and disyllabic words.

I also address a long-standing controversy concerning the tonal inventory of LT (see §2.1). I argue that LT has three lexical tones (H, LH, and L) – L being limited to some suffixes – and propose that its word tonology has been analyzed by means of two main components: prosodic phrasing defines tonal domains (prosodic words and prosodic word groups), and a set of tone rules operates in those domains. These tone rules are similar to those proposed by Duanmu (1992), but have been improved to accurately predict the tone patterns of long polysyllabic words. Also, the RLS has to be enriched to include a downstep rule that applies inside the prosodic word group. When two high tones are adjacent in a prosodic word group, the second one must be downstepped to a !H, a process that can be attributed to the Obligatory Contour Principle. Hence, I propose the following reformulation of Duanmu's *Lhasa Tone Rules* (LS):

(23) Revised Lhasa Tone Rules (RLS)

a. *Prosodic word formation*: Form prosodic words of a maximum length of two syllables, based on morphemic boundaries and semantic symmetry.

b. *Non-initial tone deletion*: Delete all non-initial tones (where LH is treated as a single unit) within a prosodic word (ω).
c. *Tone association*: Associate tones to syllables one-to-one and left-to-right within a prosodic word wherever there is an available TBU (where LH is decomposable). LH, however, can associate with a monosyllabic prosodic word (where LH is treated as a single unit).
d. *Downstep*: If two H tones are adjacent to each other across prosodic word within a prosodic word group, the second H is downstepped.

I also demonstrate through an acoustic study that there is no evidence to support the proposal that LT has a stress system realized through rhyme duration, pitch, or vocal intensity (see 3.4). These acoustic cues are only used to signal prosodic word boundaries and word-final syllables. It is more economical to analyze LT as a stressless language. I argue that the stress system that can be reconstructed for earlier stages of Tibetan is no longer synchronically active, but has been phonologically reinterpreted as a part of the tone system.

6.2 LT Intonational phonology

In Chapter 4, I argue that LT forms intonational phrases around clauses and marks them with final lengthening, pitch reset and a limited set of boundary tones (H% and L%). Although communicative functions and information structure are mostly realized by means of final particles and morphosyntactic devices in LT, I show that boundary tones, focal tones and deaccenting interact with word tones to form complex melodic patterns. The surface realizations of phrase-final boundary tones mostly agree with the *Frequency Code*, where L% communicates finality (e.g., declarative sentences) and H% communicates non-finality (e.g., continuatives and interrogative sentences without final particles) (Ohala, 1984). Boundary tones are associated to the last syllable (i.e., TBU) of the intonational phrase. As L% and H% are respectively lower and higher than corresponding lexical tones, the phrase-final L% and the tone (H or L) of the last syllable form a falling contour (i.e., H + L% or L + L%) , while the phrase-final H% and the tone of the last syllable form a rising contour (i.e., H + H% or L + H%).

In LT, a focalized constituent tends to be realized with greater vocal effort, in agreement with the *Effort Code* (Gussenhoven, 2004, p.85), where the speaker increases his or her articulation effort to optimize informational saliency. This typically results in greater amplitude, longer duration and, crucially, a raised and expanded pitch range. It seems that both lexical tones and boundary tones undergo this raising and pitch range expansion when their TBU is under focus. Interestingly, focused constituents are often followed by a span of de-emphasis, or post-focal compression (PFC), where the pitch range of the tones (especially high tones) are significantly compressed.

6.3 LT F0 Model

In Chapter 5, I present a phonologically-based F0 synthesis model to verify the adequacy of the proposed autosegmental-metrical model of LT. This F0 synthesis model consists of three main components: a) the tonal targets defined in previous chapters, b) an F0 interpolation component based on the qTA model (Xu, 2004), and c) an evaluation component allowing a comparison of the F0 contours of real LT utterances with resynthesized F0 contours of the same utterances. The tonal targets correspond to the surface form of various tones with/without the modulation effect of focus and de-emphasis (see §5.5.1). The F0 interpolation requires five inputs: a speaker-specific pitch range and declination rate, and fixed F0 targets, target approximation rates (λ), and F0 target alignments (FTA) (see §5.5.2). The F0 contours synthesized by the model correlate highly with the extracted F0 contours of real recordings of LT participants, suggesting that the current proposed phonological model adequately captures the overall tonal and intonational phonology of LT (see §5.5.3).

6.4 Future research

An indisputable deficiency of this dissertation is that it is based on only eight LT participants, including a single female speaker. Furthermore, these participants are solely from the Canadian-Tibetan community in Ottawa and Toronto. Thus, a larger body of data including more participants from various LT speech communities and a more balanced gender distribution is needed for future research. A perceptual investigation should also be integrated into future studies since there is only one known study carried out almost two decades ago (Kong, 1995). Such a study could provide further insight into the tonal and intonational phonology of LT and would allow a better tuning of the F0 resynthesis model.

It may also be necessary to compare the prosody of Canadian-Tibetan speakers with that of speakers living in Tibet and in other diasporic communities. Currently, LT only has a two-way contrastive tone system but no contrastive pitch contours (i.e., falling vs. rising) as other Tibetan tonal languages (Sun, 1997). However, Mandarin has a contrastive pitch contour shape which might sensitize bilingual LT speakers towards the pitch patterns of their language and accelerate the tonogenesis process of LT (Tournadre, 2013, cf.). Thus, it is important to monitor and document the linguistic development of LT as spoken in Tibet, as this might provide a unique opportunity to monitor contact-induced tonal developments.

The significant role of information structure in LT phrase prosody was not anticipated in my original research plan. As a result, the picture of information structure presented in this dissertation is admittedly over-simplistic. In future research, I intend to revisit LT information structure following a more systematic typological and theoretical framework allowing me to build better adapted elicitation techniques. The goal is to better control for pragmatic factors such as given information, presupposition and topicalization, and to try to formalize focus and deaccenting/post-focal compression in gradient rather than categorical terms.

The LT F0 model constructed in this dissertation is an important tool to verify the proposed phonological analyses and could be improved in terms of data processing and linguistic power. First of all, tonal targets and segments modeled are currently manually labelled in PRAAT, which is time-consuming and makes it impractical to look at a large corpus. Thus, it is imperative to develop a forced-alignment script for Lhasa Tibetan that can identify the segments and surface tones of an utterance and automatically generate labels for them. Second, the effect of micro-prosody should be integrated into the model to improve the naturalness of F0 contours. Third, there should be an interface to automatically transfer the synthesized F0 contour generated by the current model to the pitch manipulation module of PRAAT so that actual resynthesized speech can be generated and used in perception experiments, with the goal of eventually developing real LT text-to-speech tools. Lastly, it will be important to try to encode emotions into the F0 synthesis module. This is a great challenge, but it should be possible to achieve some ground work by carefully annotating a corpus of Tibetan radio or TV series.

I hope that the fundamental research on LT prosody and F0 synthesis found in this dissertation can contribute to research in LT speech synthesis and speech recognition. These technologies are quickly becoming important user interfaces to access current, and most likely future, computing devices. They have already become widely available for major languages, but there has been less economical incentive to develop them for smaller languages. This can potentially threaten the vitality of minority languages such as LT by making them less relevant in the modern world.

Bibliography

Agha, A. (1993). *Structural form and utterance context in Lhasa Tibetan : grammar and indexicality in a non-configurational language*. Monographs in linguistics and the philosophy of language,. Peter Lang, New York.

Anderson, M. D., Pierehumbert, J. B., and Liberman, M. Y. (1984). Synthesis by rule of English intonation patterns. In *Proceeding of the International Conference of Acoustics, Speech, and Signal Processing*, pages 2.8.1–2.8.4.

Barr, D. J., Levy, R., Scheepers, C., and Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, 68(3):255–278.

Bartee, E. (2000). *A beginning textbook of Lhasa Tibetan*, chapter A note on verbalizer in Lhasa Tibetan, pages 193–196. National Press for Tibetan Studies.

Beckman, M. E. and Pierehumbert, J. B. (1986). Intonational structure in English and Japanese. *Phonology Yearbook*, 3:255–310.

Bell, C. A. (1919). *Grammar of colloquial Tibetan*. The Bengal secretariat book depot, Calcutta, 2d edition.

Bielmeier, R. (1988). On tone in Tibetan. In Ueback, H. and Panglung, J. L., editors, *Tibetan Studies: Proceedings of the 4th Seminar of the International Association for*

Tibetan Studies, Munich. Kommission für Zentralasiatische Studien, Bayerische Akademie der Wissenschaften.

Bruce, G. (1977). Swedish word-accents in sentence perspertive.

Brunelle, M. (2017). Stress and phrasal prominence in tone languages: The case of Southern Vietnamese. *Journal of International Phonetic Association*, pages 1–38.

Brunelle, M., Duong, N. D., and Hùng, N. K. (2010). A laryngographic and laryngoscopic study of Northern Vietnames tones. *Phonetica*, (67):147–169.

Campbell, L. (2004). *Historical linguistics : An introduction*. MIT Press, Cambridge, Mass., 2nd edition. Lyle Campbell. ill., maps ; 24 cm.

Cao, Y. (1930). A system of tone letters. *Le Maitre Phonetique*, (45):24–27.

Caplow, N. J. (2009). *The role of stress in Tibetan tonogenesis: A study in historical comparative acoustics*. Thesis.

Caplow, N. J. (2016). Epistemics and evidentials in diasporic common tibetan. To appear in Linguistics of the Tibeto-Burman Area.

Chang, K. and Chang, B. S. (1964). *A manual of spoken Tibetan (Lhasa dialect)*. University of Washington Press, Seattle,.

Chen, S., Wang, B., and Xu, Y. (2009). Closely related languages, different ways of realizing focus. In *Proceedings of Interspeech 2009*, Brighton, UK.

Chen Qi, Yu Hongzhi, C. C. S. J. (2010). A study on Tibetan prosodic model of speech and respiratory signals. In *Proceedings of the 2010 IEEE International Conference on Information and Automation*, Harbin, China.

Civera, M. B. (1970). *The distinctive phonology of Lhasa Central Tibetan: Results of a Field Trip to study the Himalayas as an Ethno-linguistics Area*. Thesis.

Cruttenden, A. (1997). *Intonation*. Cambridge textbooks in linguistics. Cambridge University Press, Cambridge U.K. ; New York, NY, USA, 2nd edition.

Dawson, W. (1980). *Tibetan Phonology*. Ph.D. Thesis.

DeLancey, S. (1990). Ergativity and the cognitive model of event structure in Lhasa Tibetan. *Cognitive Linguistics*, 1(3):289–321.

DeLancey, S. (1991). The origins of verb serialization in modern Tibetan. *Studies in Language*, 15(1):1–23.

DeLancey, S. (1992). The historical status of the conjunct/disjunct pattern in Tibeto-Burman. *Acta Linguistica Hafniensia*, 25:39–62.

DeLancey, S. (2003a). Classical Tibetan. In Thurgood, G. and Lapolla, R. J., editors, *The Sino-Tibetan Languages*, pages 255–269. Routlege, London and New York.

DeLancey, S. (2003b). Lhasa Tibetan. In Thurgood, G. and Lapolla, R. J., editors, *The Sino-Tibetan Languages*, pages 270–288. Routlege, London and New York.

Denwood, P. (1999). *Tibetan*. London Oriental and African language library,. John Benjamins Pub. Co., Amsterdam ; Philadelphia.

Desideri, F. I. (2010). *Mission to Tibet: The Extraordinary Eighteenth-Century Account of Father Ippolito Desideri S. J.* Wisdom Publication.

Dotson, B. (2010). *The Old Tibetan Annals: An Annotated Translation of Tibet's First History, With an Annotated Cartographical Documentation by Guntram Hazod*. Verlag der Österreichischen Akademie der Wissenschaften, Vienna.

Duanmu, S. (1990). *A formal study of syllable, tone, stress and domain in Chinese languages*. Thesis.

Duanmu, S. (1992). An autosegmental analysis of tone in four Tibetan languages. *Linguistics of the Tibeto-Burman Area*, 15(1).

Duanmu, S. (2002). *The Phonology of Standard Chinese*. Oxford University Press, New York.

Duff, T. (2005). Standard Tibetan grammar: Vol 1: The thirty verses of Minister Thumi.

Duff, T. (2015). *The Illuminator Tibetan-English Encylopædic Dictionary*.

Fang Hu, Z. X. (2010). Lhasa tones. In *Speech Prosody 2010 Conference Proceedings*, Chicago, Illinois.

Ferlus, M. (2004). The origin of tones in Viet-Muong. In Burusphat, S., editor, *Eleventh Annual Conference of the Southeast Asian Linguistics Society 2001,*, pages 297–313, Tempe, Arizona. Arizona State University Programme for Southeast Asian Studies Monograph Series Press.

Fletcher, J. (2013). The prosody of speech: Timing and rhythm. In Hardcastle, W. J., Laver, J., and Gibbon, F. E., editors, *The handbook of phonetic sciences*, Blackwell Handbooks in Linguistics, chapter 15. Blackwell Publishing Ltd.

Fujisaki, H. and Nagashima, S. (1969). A model for the synthesis of pitch contours. *Annual Report of the Engineering Research Institute*, 28:53–60.

Garrett, E. J. (2001). *Evidentiality and Assertion in Tibetan*. Thesis.

Garrett, E. J. and Hill, N. W. (2015). Constituent order in the tibetan noun phrase. *SOAS Working Papers in Linguistics*, 17:35–48.

Goldsmith, J. A. (1976). *Autosegmental Phonology*. PhD thesis, MIT.

Goldsmith, J. A. (1980). Review of tone: A linguistic survey. *Language*, 56(2):413–418.

Goldstein, M. C. and Narkyid, N. (1984). *English-Tibetan dictionary of modern Tibetan*. University of California Press, Berkeley. 83018119 compiled by Melvyn C. Goldstein with Ngawangthondup Narkyid. 28 cm. In English.

Goldstein, M. C. and Nornang, N. (1970). *Modern spoken Tibetan: Lhasa dialect*. University of Washington Press, Seattle,. 68013412 by Melvyn C. Goldstein and Nawang Nornang. 29 cm.

Goldstein, M. C., Rimpoche, G., and Phuntshog, L. (1991). *Essentials of modern literary Tibetan : A reading course and reference grammar*. University of California Press, Berkeley.

Goldstein, M. C., Shelling, T. N., and Surkhang, J. T. (2001). *The new Tibetan-English dictionary of modern Tibetan*. University of California Press, Berkeley.

Gong, X. (2014). Structure of the prosodic word and stress in Old Tibetan. In *SEALS24*, Yangon University.

Gu, C. (2014). Smoothing spline ANOVA models: R package GSS. *Journal of Statistical Software*, 58(5):1 –25.

Guo, S., Gao, L., and Yu, H. (2011). Research on Lhasa Tibetan prosodic model of journalese based on respiratory signal. In *2011 International Conference on Image Analysis and Signal Processing (IASP)*, Hubei, China.

Gussenhoven, C. (1984). *On the grammar and semantics of sentence accents*. Publications in language sciences ; 16. Foris Publications, Dordrecht, Holland ; Cinnaminson, N.J., U.S.A.

Gussenhoven, C. (2002). Intonation and interpretation: Phonetics and phonology. In Bel, B. and Marlien, I., editors, *Proceedings of the speech prosody*, pages 47–57, Aix-en Provence. Université de Provence.

Gussenhoven, C. (2004). *The phonology of tone and intonation*. Research surveys in linguistics. Cambridge University Press, Cambridge ; New York. Carlos Gussenhoven. ill. ; 24 cm.

Hall, T. A. and Hildebrandt, K. A. (2008). Phonological and morphological domains in Kyirong Tibetan. *Linguistics*, 46(2):215–248.

Hannah, H. B. (1912). *A grammar of the Tibetan language, literary and colloquial.* Printed at the Baptist mission press, Calcutta,.

Hari, A. M. (1977). *An Investigation of the tones of Lhasa Tibetan*, volume 13 of *Asian-Pacific Series*. SIL, Huntington Beach, CA.

Hartmann, K. (2007). Focus and tone. *Interdisciplinary Studies on Information Structure*, 6:221–235.

Haudricourt, A.-G. (1954). De l'origine des tons en vietnamien. *Journal Asiatique*, 242:69–82.

Haudricourt, A.-G. (1961). Bipartition et tripartition des systèmes de tons dans quelques langues d'extrême-orient. *Bulletin de la Societe de Linguistique de Paris*, 56:163–180.

Hayes, B. (1989a). Compensatory lengthening in moraic phonology. *Linguistic Inquiry*, 20(2):253–306.

Hayes, B. (1989b). The prosodic heirarchy in meter. *Phonetics and phonology*, 1:201–260.

Hayes, B. P. (1976). *A Metrical theory of stress rules*. PhD thesis, MIT.

Hayes, B. P. (1995). *Metrical Stress Theory:Principles and Case Studies*. University of Chicago Press, Chicago.

Hetenyi, E. (1972). Alexander Csoma De Koros: The Hungarian Bodhisattva. *Bulletin of Tibetology*, IX(1):34–41.

Hildebrandt, K. A., Bickel, B., and Schiering, R. (2010). The prosodic hierarchy is not universal. *Journal of Linguistics*, 46:657–709.

Hill, N. W. (2010). An overview of Old Tibetan synchronic phonology. *Transactions of the Philological Society*, 108(2):110–125.

Hirst, D. (2005). Form and function in the representation of speech prosody. *Speech Communication*, 46:334–347.

Hirst, D. and de Cristo, A. (1998). *Intonational Systems: A Survey of Twenty Languages*. Cambridge University Press, Cambridge.

Hirst, D. and Espesser, R. (1993). Automatic modelling of fundamental frequency using a quadratic spine function. *Travaux de I'Institute de Phonétique D' Aix*, 15:75–85.

Hombert, J.-M., Ohala, J. J., and Ewan, W. G. (1979). Phonetic explanations for the development of tones. *Language*, 55(1):37–37–58.

Hoshi, I. (2003). *A verb dictionary of the Modern Spoken Tibetan of Lhasa*. Number 42 in Asia and African Lexicon Series. Research Institute for the Lnaguages adn Cultures of Asia and Africa, Tokyo.

Hu, T. (1980). Zanyu(lasa hua) de shengdiao yanjiu [research on the tones of Lhasa Tibetan]. *Minzu Yuwen*, 1:22–36.

Hu, T. (1988). A comparative study of tonal and toneless Tibetan dialects. *Languages and history in East Asia; festschrift for Tatsuo Nishida on the occasion of his 60th birthday*, pages 75–92.

Hu, T. (2002). *Zang yu yan jiu wen lun (Research of Tibetan languages)*. Tibetology of China.

Hu, T., Au, A., and Lin, L. (1982). Zangyu lhas hua shengdiao shiyan [experiments in Lhasa Tibetan tone]. *Yuyan Yanjiu*, pages 18–38.

Huang, B. (1986). 12-13 century Tibetan (Central Tibetan): A preliminary investingation in phonology. *Minzu Yuwen*, (3).

Huang, X. D., Ariki, Y., and Jack, M. A. (1990). *Hidden Markov models for speech recognition*. Edinburgh University Press.

Hyman, L. M. (1976). Phonologication. In Juilland, A., editor, *Linguistic studies offered to Joseph Greenberg*, volume 2, chapter 54, pages 407–18. Anma Libri.

Hyman, L. M. (1989). *A Theory of phonological weight*. Foris, Dordrecht, Holland.

Hyman, L. M. (2006). Word-prosodic typology. *Phonology*, 23(2):225–258.

Hyman, L. M. (2013). Towards a typology of tone system changes. National Institute for Japanese Language and Linguistics. International Conference on Phonetics and Phonology.

Hyman, L. M. and Magaji, D. J. (1970). *Essentials of Gwari grammar*. Institute of African Studies, University of Ibadan, [Ibadan].

Hyslop, G. (2009). Kurtop tone: A tonogenetic case study. *Lingua*, 119(6):827–827–845.

Ito, J. and Mester, A. (2008). The extended prosodic word. In Grijzenhout, J. and Kabak, B., editors, *Phonological domains: universals and deviations*, pages 135–194. Mouton de Gruyter, Berlin.

Jäschke, H. A. (1881). *A Tibetan-English dictionary, with special reference to the prevailing dialects*. Printed by Unger brother (T. Grimm), London Berlin,.

Jäschke, H. A. (1883). *Tibetan*, volume VII of *Trübner's collection of Simplified Grammars of the principal Asiatic and European Languages*. Trübner and Co, London.

Jun, S.-A. (2014). Prosodic typology: By prominence type, word prosody, and macro-rhythm. In Jun, S.-A., editor, *Prosodic typology II: The phonology of intonation and phrasing*, Oxford linguistics, chapter 17., pages 520–539. Oxford University Press.

Kager, R. (2006). Feet and metrical stress. In de Lacey, P., editor, *The Cambridge Handbook of Phonology*, pages 195–227. Cambridge University Press.

Kingston, J. (2005). The phonetics of Athabaskan tonogenesis. *ATHABASKAN PROSODY, Hargus, Sharon, and Rice, Keren [Eds], Amsterdam: John Benjamins, 2005, pp 137-184*.

Kingston, J. (2011). Tonogenesis. In M. van Oostendorp, C. J. Ewen, E. H. and Rice, K., editors, *Blackwell Companion to Phonology*, volume 4, chapter 97. Blackwell Publising, Oxford, UK.

Kitamura, H. (1974). *Gendai tibetto-go no hatuon [The pronunciation of modern Tibetan]*. Institute for the Study of the Languages and Cultures of Asia and Africa, Tokyo:.

Kjellin, O. (1974). A preliminary report on Tibetan tones. *Annual bulletin, RILP*, 8:115–123.

Kochanski, G. and Shih, C. (2003). Prosody modeling with soft templates. *Speech Communication*, 39:311–352.

Koerber, H. N. v. (1935). *Morphology of the Tibetan language: A contribution to comparative Indosinology*. Suttonhouse, Los Angeles.

Kong, J. (1995). The tonal perception study of Lhasa Tibetan tone. *Minzu Yuwen*, 3(56-64).

Kong, J. (2012). A study of the origin of Tibetan tones by hononym rate.

Koros, A. C. d. (1834a). *A grammar of the Tibetan language, in English.* Printed at the Baptist mission press, Calcutta.

Koros, A. C. d. (1834b). The tibetan english dictionary.

Ladd, D. R. (1996). *Intonational phonology*. Cambridge studies in linguistics. Cambridge University Press, Cambridge ; New York.

Ladefoged, P. (2003). *Phonetic data analysis : An introduction to fieldwork and instrumental techniques*. Blackwell Pub., Malden, MA.

Lama, D. (1997). *The Four Noble Truths*. Thorsons.

Lambrecht, K. (1994). *Information structure and sentence form: Topic, focus, and the mental representations of discourse referents*. Cambridge studies in linguistics. Cambridge University Press, Cambridge ; New York, NY, USA.

Lee, A. and Xu, Y. (2012). Revisiting focus prosody in Japanese. In *Proceedings of Speech Prosody 2012*, pages 274–277, Shanghai.

Lee, Y. and Xu, Y. (2010). Phonetic realization of contrastive focus in Korean. In *Proceedings of Speech Prosody 2010*, pages 100033:1–4, Chicago.

Lewis, M. P., Simon, G. F., and Fenning, C. D., editors (2014). *Ethnologue: Lanugages of the World*. SIL International, Dallas, TX, seventeenth edition edition.

Liberman, M. (1975). *The intonational system of English*. PhD thesis, MIT.

Lim, J. K. S. (2010a). The pitch realiasation of genitive phrase in Lhasa Tibetan. Paper present at the Montreal-Ottawa-Toronto Phonology Workshop, Carleton University.

Lim, J. K. S. (2010b). Voicing contrast of simple stops in Lhasa Tibetan.

Lim, J. K. S. (2012). The pitch realization and tonal phonology of Lhasa Tibetan. Poster presented at the 5th International Conference on Tone and Intonation in Europe (TIE5), Oxford University.

Lim, J. K. S. (2013). The tonal phonology and pitch realization of polysyllabic words in Lhasa Tibetan. Paper presented at the 46th International Conference on Sino-Tibetan Languages and Linguistics (ICSTLL46), Dartmouth College.

Lim, J. K. S. (2014). The tonal and intonational phonology of Lhasa Tibetan. The 6th International Conference on Tone and Intonation in Europe (TIE6), Utrech University.

Lim, J. K. S. (2016). Prosodic structure and word tonology of Lhasa Tibetan. The 7th Tone and Intonational in Europe (TIE7), University of Kent.

Lu, G., Hongzhi, Y., Zhang, J., and Fang, H. (2011). Research on HMM based speech synthesis for Lhasa dialect. In *2011 International Conference on Image Analysis and Signal Processing*, pages 429–433.

Lu, G., Yu, H., Zhang, J., and Fang, H. (2013). HMM-based Tibetan Lhasa speech synthesis system. In *2013 3rd International Conference on Computer Science and Network Technology (ICCSNT)*, pages 92–95.

Maspero, H. (1912). Etude sur la phonétique historique de la langue annamite. les initiale. *Bulletin de l'Ecôle Française d'Extême-Orient*, 12(1):127.

Matisoff, J. A. (1973). Tonogenesis in Southeast Asia. In Hyman, L. M., editor, *Consonant types and tones*, pages 71–95. The Linguistic Program, University of Southern California.

Matisoff, J. A. (2000). Tibetan-Burma tonology in an areal context. In *The Fifth International Symposium on Languages and Linguistics*, pages 85–151, Ho Chi Minh City. Vietnam National University, Ho Chi Minh City University of Social Science and Humanities.

Matisoff, J. A. (2003). *Handbook of Proto-Tibeto-Burman: System and Philosophy of Sino-Tibetan Reconstruction*. University of California Press, Berkeley, Los Angeles, London.

Mazaudon, M. (1977). *Tibeto-Burman Tonogenetics*. PhD thesis, Dept. of Linguistics, University of California Berkeley,, California.

McCarthy, J. and Prince, A. (1993). Prosodic morphonology I: Constraint interaction and satisfaction. *Linguistics Department Faculty Publication Series*, (14).

McCarthy, J. J. (1986). OCP effects: Gemination and antigemination. *Linguistic Inquiry*, 17:207–263.

Meredith, S. (1990). *Issues in the phonology of prominence*. Thesis.

Michaud, A. and Brunelle, M. (2016). Information structure in Asia: Yongning Na (Sino-Tibetan) and Vietnamese (Austroasiatic). In Fery, C., editor, *The Oxford Handbook of Information Structure*, chapter 38. Oxford University Press.

Michaud, A. and Vaissière, J. (2015). Tone and intonation: Introductory notes and pratical recommendation.

Miller, R. (1955). Studies in spoken Tibetan 1: Phonemics. *Journal of the American Oriental Society*, 75:46–51.

Monich, I. (2011). Interaction of tone and intonation in Lhasa Tibetan: A working hypothesis. In *Generative Linguistics in the Old World 34*, Wien. Department of Linguistics, Universität Wien.

Nespor, M. and Vogel, I. (1982). Prosodic domains of external sandhi rules. In van der Hulst, H. and Smith, N., editors, *The structure of Phonological representation*, number 1, pages 225–255. Foris Publications, Dordrecht, Holland.

Nespor, M. and Vogel, I. (1984). *Prosodic Phonology*. Foris Publications, Dordrecht, Holland.

Nespor, M. and Vogel, I. (2007). *Prosodic Phonology*, volume 28 of *Studies in Generative Grammar*. Mouton de Gruyter, Berlin New York.

Nishida, T. (1970). *Chibetto gengogaku josetsu "A studies on sixteen century Tianquan dialect of Tibetan languages in West Kham Provice"*. Seibankan yakugo no kenkyū. Shokado, Kyota.

Nishida, T. (1975). On the development of tones in Tibetan. *Acta Asiatica*, 29:43–55.

Odden, D. (1979). Principles of stress assignment: A crosslinguistics view. *Studies in the Linguistic Sciences*, 9(1):157–175.

Odden, D. (1988). Antigemination and the OCP. *Linguistic Inquiry*, 19:451–475.

Ohala, J. J. (1984). An ethological perspective on common cross-language utilization of f0 of voice. *Phonetica*, 41:1–16.

Öhman, S. and Lindqvist-Gauffin, J. (1965). Analysis-by-synthesis of prosodic pitch contours. *Quarterly Progress and Status Report - Royal Institute of Technology, Department of Speech, Music and Hearing*, 6(4):1–6.

Ossorio, J. R. W. (1982). *Tsang Tibetan Phonology*. Thesis.

Peperkamp, S. (1999). Prosodic words. *GLOT International*, pages 15–16.

Pierrehumbert, J. (1980). *The phonology and phonetics of English intonation*. Thesis.

Pierrehumbert, J. (1981). Synthesis of intonation. *Journal of Acoustical Society of America*, 70(4):985–995.

Pierrehumbert, J. B. and Beckman, M. E. (1988). *Japanese tone structure*. Linguistic inquiry monographs 15. MIT Press, Cambridge, Mass.

Pomplum, T. (2010). *Jesuit on the Roof of the World: Ippolito Desideri's Mission to Tibet*. Oxford University Press, New York.

Prom-on, S., Xu, Y., and Thipakorn, B. (2009). Modeling tone and intonation in Mandarin and English as a process of target approximation. *Journal of Acoustical Society of America*, 125(1):405–424.

Purcell, E. T. (1974). From dynamic stress to word tone: The consonantal origins of Balto-Slavic tone. *Papers from the Regional Meetings, Chicago Linguistic Society*, 10(0577-7240):528–528–538.

Purcell, E. T., Villegas, G., and Young, S. P. (1978). A before and after for tonogenesis. *Phonetica*, 35(5):284–284–293.

Qu, A. (1981). Zanyu de biandian (Tibetan tone sandhi). *Minzu Yuwen*, 81(4):20–26.

Qu, A. and Tan, K. (1983). *Ali Zangyu [Ali Tibetan], zhongguo shehui kexue yan chubanshe*. Chinese Academy of Social Sciences Press, Bejing.

Raska, J. (2013). Tibetan immigration to Canada.

Remijsen, B. (2002). Lexically contrastive stress accent and lexical tone in ma‘ya. In Warner, C. G. . N., editor, *Laboratory Phonology VII*, pages 585–614, Berlin/New York. Mouton de Gruyter,.

Riad, T. (2013). *The phonology of Swedish*. The phonology of the world's languages. Oxford University Press.

Rogers, H. (2005). *Writing systems: A linguistic approach*, volume 18. Blackwell Pub., Malden, MA.

Rörich, G. d. (1931). Modern Tibetan phonetics, with special reference to the dialect of Central Tibet. *JRSB*, 27.

Rose, P. (1993). A linguistic-phonetic acoustic analysis of Shanghai tones. *Australina Journal of Linguisticss*, (13):185–220.

Rose, P. (2009). Singing the syllables: translating spelling into music in Tibetan spelling chant. In Chan, A. and Noble, A., editors, *Sounds in translation: intersections of music, technology and society*, pages 145–167. ANU ePress, Canberra Australia.

Ross, K. and Ostendorf, M. (1999). A dynamical system model for generating fundamental frequency for speech synthesis. *IEEE Trans. Speech and Audio Proc.*, 7:295–309.

Sandberg, G. (1894). *Handbook of Colloquial Tibetan: A practical guide to the language of Central Tibet.* Calcutta.

Sedláček, K. (1959). The tonal system of Tibetan (Lhasa Dialect). *T'oung Pao*, 47(3/5):181–250.

Selkirk, E. (1995). The prosodic structure of function words. In *University of Massachusetts occasional papers 18: Papers in Optimality Theory*, pages 439–469. GLSA, University of Massachusetts, Amherst.

Selkirk, E. O. (1980). Prosodic domains in phonology: Sanskrit revisited. In Aronoff, M. and Kean, M., editors, *Juncture*, pages 107–129. Anma Libri, California.

Selkirk, E. O. (1984). *Phonology and syntax.* MIT Press, Cambridge, MA.

Selkirk, E. O. (2002). The syntax-phonology interface. In Comrie, B., editor, *International Encyclopedia of the Social and Behavioral Sciences*, chapter 3.9/23. Elsevier.

Selkirk, E. O. (2011). The syntax-phonology interface. In John Goldsmith, Jason Riggle, A. C. L. Y., editor, *The Handbook of Phonological Theory*, Blackwell Handbooks in Linguistics. Wiley-Blackwell, 2nd edition.

Shafer, R. (1968). *Introduction to Sino-Tibetan: Part 1*. Otto Harrassowitz, Wiesbaden.

Shih, C. (2000). A declination model of Mandarin Chinese. In Botinis, A., editor, *Intonation: Analysis, Modelling and Technology*. Kluwer Academic Publisher, NY.

Simons, F. G. and Fenning, C. D., editors (2017). *Ethnologue: Languages of the World*. SIL International, twentieth edition.

Sprigg, R. K. (1955). The tonal system of Tibetan (Lhasa Dialect) and the nominal phrase. *Bulletin o fthe School of oriental and African Studies*, 17(1):133–153.

Sprigg, R. K. (1972). A polysystemic approach, in Proto-Tibetan reconstruction, to tone and syllable-initial consonant clusters. *Bulletin of the School of Oriental and African Studies*, 35(3):546–587.

Sprigg, R. K. (1993). Controversy in the tonal analysis of Tibetan. *Bulletin of the School of Oriental and African Studies, University of London*, 56(3):470–501.

Sugahara, M. (2003). *Downtrends and post-FOCUS intonation in Tokyo Japanese*. PhD thesis, Uiversity of Massachusetts, Amherst.

Sun, J. (1997). The typology of Tibetan tone. In *Symposium Series of the Institute of History and Philology*, Taipei.

Sun, J. (2003). Variegated tonal developments in Tibetan. In Bradley, D., LaPolla, R., Michailovsky, B., and Thurgood, G., editors, *Language Variation: Papers on* *variation and change in the Sinosphere and in the Indosphere in honour of James A. Matisoff*, chapter 3. Canberra: Pacific Linguistics.

Sun, X. (2002). *The determination, analysis, and systhesis of fundamentala frequency*. PhD thesis, Northwestern University.

Svantesson, J.-O. (1989). Tonogenetic mechanisms in Northern Mon-Khmer. *Phonetica*, 46:60–79.

't Hart, J., Collier, R., and Cohen, A. (1990). *A perceptual study of intonation: An experimental-phonetic approach*. Cambridge University Press.

Taheri Ardali, M. and Xu, Y. (2012). Phonetic realization of prosodic focus in Persian. In *Proceedings of Speech Prosody 2012*, pages 326–329, Shanghai.

Takeuchi, T. (2012). Old Tibetan buddhist texts from the Post-Tibetan Imperial Period (mid-9 c. to late 10 c.). In Scherrer-Schaub, C., editor, *Old Tibetan Studies dedicated to the memory of R.E. Emmerick*. Brill.

Taylor, P. (2000). Analysis and synthesis of intonation using the tilt model. *Journal of Acoustical Society of America*, pages 1697–1714.

Thonden, L. (1986). *Moden Tibetan Language*, volume I. Library of Tibetan Works and Archives, Dharamsala.

Thurgood, G. (2002). Vietnamese and tonogenesis: Revising the model and the analysis. *Diachronica*, 19(2):333–363.

Thurgood, G. and LaPolla, R. J. (2003). *The Sino-Tibetan Languages*. Routledge, London ; New York.

Tournadre, N. (1991). The rhetorical use of the Tibetan ergative. *Linguistics of the Tibeto-Burma Area*, 14(1):93–108.

Tournadre, N. (1996). *L'ergativité en tibétain : Approche morphosyntaxique de la langue parlée*. Bibliothèque de l'information grammaticale. Peeters, Louvain ; Paris. 97151841 Nicolas Tournadre. ill., map ; 24 cm. Includes bibliographical references (p. [377]-388).

Tournadre, N. (2013). *Trans-Himalayan linguistics: historical and descriptive linguistics of the Himalayan area*, volume 266 of *Trends in linguistics. Studies and monographs*. De Gruyter Mouton.

Tournadre, N. and Dorje, S. (2003). *Manual of Standard Tibetan : Language and civilization : Introduction to Standard Tibetan (spoken and written) followed by an appendix on classical literary Tibetan*. Snow Lion Publications, Ithaca, N.Y.

Truckenbrodt, H. (2007). The syntax-phonology interface. In de Lacey, P., editor, *The Cambridge Handbook of Phonology*, chapter 18. Cambridge University Press, Cambridge, UK.

Utsumi, A. (2016). The intonation of interrogatives and focused elements in the Bantik language. In *The third International workshop on Information structure of Austronesian Languages*, pages 18–20, Tokyo.

van der Hulst, H. (2012). Deconstructing stress. *Lingua*, 122:1491–1521.

van Schaik, S. (2010). Brandon Dotson: The Old Tibetan Annals: An annotated translation of Tibet's first history, with an annotated cartographical documentation by Guntram Hazod. *European Bulletin of Himalayan Research*, 37:149–152.

Vigário, M. (2010). Prosodic structure between the prosodic word and the phonological phrase: Recursive nodes or an independent domain? *The Linguistic Review*, 27:485–530.

Vigário, M. C. (2003). *The prosodic word in European Portuguese*. Mouton de Gruyter, Berlin and New York.

Wang, B., Wang, L., and Qadir, T. (2011). Prosodic realization of focus in six languages/dialects in China. In *Proceedings of the 17th International Congress of Phonetic Sciences*, pages 17–21, Hong Kong.

Wang, L., Wang, B., and Xu, Y. (2012). Prosodic encoding and perception of focus in Tibetan (Anduo dialtect). In *Proceedings of Speech Prosody 2012*, pages 286–289, Shanghai.

Weenink, P. B. and David (2017). Praat: Doing phonetics by computer (version 6.0.29).

Wylie, T. (1959). A standard system of Tibetan transcription. *Harvard Journal of Asiatic Studies*, 22:261–267.

Xu, L. (2004). Manifestation of informational focus. *Lingua*, 114(3):277–299. Lingua March 20, 2007.

Xu, Y. (2005). Speech melody as articulatorily implemented communicative functions. *Speech Communication*, 46:220–251.

Xu, Y. (2011). Post-focus compression: Cross-linguistic distribution and historical origin. In *Proceedings of the 17th International Congress of Phonetic Sciences*, Hong Kong.

Xu, Y. and Wang, E. (2001). Pitch targets and their realization: Evidence from Mandarin Chinese. *Speech Communication*, (33):319–337.

Xu, Y. and Xu, C. X. (2005). Phonetic realization of focus in English declarative intonation. *Journal of Phonetics*, 33:159–197.

Xu, Y., Xu, C. X., and Sun, X. (2004). On the temporal domain of focus. In *Proceedings of Internation Conference on Speech Prosody*, pages 81–84, Nara, Japan.

Yip, M. (1988). The obligatory contour principle and phonological rules: A loss of identity. *Linguistic Inquiry*, 19(1):65–100.

Yip, M. J. W. (1993). Tonal register in east asian languages. In van der Hulst, H. and Snider, K., editors, *The phonology of Tone*, number 17 in Linguistic models, pages 245–268. Mouton de Gruyter.

Yip, M. J. W. (2002). *Tone*. Cambridge textbooks in linguistics. Cambridge University Press, Cambridge ; New York. Moira Yip. maps ; 24 cm.

Yu, D. and Chao, Y. (1930). *Love songs of the Sixth Dalai Lama, Tshangs-dbyangsɹgya-mtsho*. Series A. Academia Sinica, Peiping.

Yu, K. and Young, S. (2010). Continuous f0 modeling for HMM based statistical parametric speech synthesis. *IEEE*, 19(5).

Zec, D. (2007). The syllable. In de Lacey, P., editor, *The Cambridge Handbook of Phonology*, pages 161–191. Cambridge University Press, Cambridge, UK.

Zerbian, S., Genzel, S., and Kügles, F. (2010). Experimental work on prosodically-marked information structure in selected African languages (Afroasiatic and Niger-Congo). In *Proceedings of Speech Prosody 2010*, pages 1–4, Chicago.

Appendix A

Morphology of LT

1.1 Derived nouns

Stem			Suffix		Derived noun	
bod	/pʰø̌ʔ/	'Tibet'	*pa*	/-pá/	[pʰø̀ʔpa]	'Tibetan'
khang	/kʰáŋ/	'building'	*pa*	/-pá/	[kʰáŋ]	'house'
dar	/tʰǎr/	'silk'	*cha*	/-tɕá/	[tʰàrtɕá]	'flag'
skar	/káː/	'astrology'	*ma*	/-má/	[káːmá]	'star'
las	/lěː/	'to do (*fr. Sanskrit*)'	*ka*	/-ká/	[lèːká]	'work'
rtags	/tʰé/	'sign'	*tsi*	/-tsí/	[tʰétsí]	'emblem'
re	/rě/	'want for something'	*ba*	/-wá/	[rèwá]	'hope'

Table A.1: Derived nouns of LT :stem-suffix (Qu and Tan, 1983).

1.2 Possible combinations of grammatical categories in LT disyllabic compound nouns

Denwood (1999, p.91) lists all known combination of grammatical categories in disyllabic compound nouns; and his examples are reproduced in Table A.2.

Common				
noun + noun	*mar phyur*	[mar.ɕur]	(butter.cheese)	'dairy product'
noun + adjective	*stod thung*	[tyː.døː]	(trunk.short)	'shirt'
adjective + noun	*dkar rtsi*	[kaː.dziː]	(white.resin)	'whitewash'
Less common				
adjective + adjective	*ring thung*	[riŋ.tʰuŋ]	(long.short)	'length'
verb + verb	*'gro song*	[dro.søː]	(go.went)	'expenses'
verb + noun	*rgyugs.hyi*	[kʲug.kʰi]	(run.dog)	'running dog'
noun + verb	*ngo sprod*	[ŋo.drøː]	(face.join)	'introduction'
adjective + verb	*gsang gcod*	[saŋ.tɕoː]	(secret.cut off)	'bathroom'

Table A.2: Possible combinations of grammatical categories in LT disyllabic compound nouns.

1.3 Verbalized nouns

In Table A.3, I partially reproduce Thonden (1986, p.1-12) extended list of verbalized nouns. Bartee (2000, p.192-193) also has a detailed note on common verbalizers in LT.

Noun			Verbalizer		
ki	[kí]	'a yell'	+*rkyab*	[kʲà]	'to yell'
dka' las	[káʔlɛʔ]	'hardship'	+*rkyab*	[kʲǎ]	'to work hard'
khebs	[kʰɛ́ː]	'a lid'	+*rkyab*	[kʲà]	'to cover'
skad	[kǿ]	'language, sound'	+*gdang*	[táŋ]	'to call, to invite'
kha	[kʰá]	'mouth'	+*gdang*	[táŋ]	'to gossip'
dshas	[ʃǎ]	'song'	+*gdang*	[táŋ]	'to sing'
skad cha	[kɛ́dtɕʰá]	'disucssion, talk'	+*ched*	[tɕʰèʔ]	'to detain'
kha ta	[kʰátá]	'guidance'	+*ched*	[tɕʰèʔ]	'to guide'
chos	[tɕʰǿ]	'dharma, religion'	+*ched*	[tɕʰèʔ]	'to practice religion'

Table A.3: Example of verbalized nouns of LT.

1.4 Secondary Verbs of serial verbs

There are two types of secondary verbs: modal verbs and tense-aspect verbs. Table A.4 is a collection of secondary verbs from various authors (DeLancey, 2003b; Garrett, 2001).

At first glance, the secondary verb might appear to occupy the same morphological position as the verbalizer or TAM suffixes. DeLancey (1991) proposes that verb serialization is the source of verbalizers and TAM suffixes in LT. Secondary verbs are different from verbalizers. Distinction between secondary verbs and TAM suffixes is equivocal, in that some secondary verbs can be used as primary verbs while others that have advanced grammaticalization are considered as TAM suffixes in LT. TAM suffixes are obligatory and come after the serial verb construction when there is one. A collection of secondary verbs can be seen in Table A.4.

Modal	*dgos*	[kǒ]	‘to need, want, have to’
	thub	[tʰúp]	‘to be able’
	shes	[ʃé]	‘to know’
	nus	[ny̌]	‘to dare’
	’dod	[tø̌]	‘to want, wish’
	snying ’dod	[ɲíŋtǿ]	‘to wish’
	bsam	[sám]	‘to think’
	gzhi	[sí]	‘to intend, plan’
Tense-aspect	*tshar*	[tsʰár]	‘to finish’
	bsdad	[tɛ̌]	‘stay’
	bzhag	[tɕǎʔ]	‘to put, to leave’
	byung	[tɕúŋ]	‘to get’
	ran	[rɛ̌n]	‘to be time to’
	’gros	[ʈǒː]	‘to be about to’
	chog	[tɕʰóʔ]	‘to be ready to’
	long	[lǒŋ]	‘to have the time to’
	yong	[jǒŋ]	‘to come’
	myong	[ɲǒŋ]	‘to taste, to experience’
	’gro	[ʈǒ]	‘to go’
	song	[sóŋ]	‘to go’

Table A.4: Examples of secondary verb.

1.5 The negative forms of main TAM suffixes in LT

	Evidential		
Tense/aspect	EGO	Direct	Indirect
Future	$V_{(pres)}$ [-ki-mɛn] $V_{(pres)}$-*ki-man*		$V_{(pres)}$ [-ki-mareʔ] $V_{(pres)}$-*ki-ma-red*
Present	$V_{(pres)}$ [-ki-meʔ] $V_{(pres)}$-*ki-med*	$V_{(pres)}$ [-ki-min-duʔ] $V_{(pres)}$-*ki-min-’dug*	$V_{(pres)}$ [-ki-joːmareʔ] $V_{(pres)}$-*ki-yod ma-red*
Imperfective Past	$V_{(past)}$ [-pa-mɛn] $V_{(past)}$-*pa-man*	$V_{(past)}$ [-ma-soŋ] $V_{(past)}$-*ma-song*	[ma-] $V_{(past)}$ [-pa-reʔ] *ma*-$V_{(past)}$-*pa-red*
Perfective Past	$V_{(past)}$ [-meʔ] $V_{(past)}$-*med*	$V_{(past)}$ [-min-duʔ] $V_{(past)}$-*min-’dug*	$V_{(past)}$ [-joːmareʔ] $V_{(past)}$-*yod-ma-red*

Table A.5: The negative forms of main TAM suffixes in LT.

Appendix B

Previous work in LT tonal phonology

Qty		Tones	Remarks and references
2	a.	high, low	SF (Jäschke, 1881)
	b.	high (falling) tone, low (rising circumflex)	SF (Yu and Chao, 1930)
	c.	/L, H/	Based on CT prefix and initials, UL (Miller, 1955, p.46)
	d.	Tone 1(high pitch level),	Tone template (Sprigg, 1955, 1993; Denwood, 1999)
		Tone 2 (low pitch level)	
	e.	Tone 1 [H(H)]	Four tones based on two features (H and L)(Kjellin, 1974, p.119)
		Tone 2 [L(H)]	
		Tone 3 [HL]	
		Tone 4 [L∅L]	
	f.	High, Low	(Ossorio, 1982, p.86)
	g.	/H, LH/	Autosegmental UF (Duanmu, 1990)
	h.	/H, LH/	OT analysis UF (Yip, 2002)
4	a.	44, 41, 13, 35	SF (Nishida, 1975)
	b.	55, 53, 12, 14	SF (Qu, 1981, p. 22)
	c.	low reg. (breathy): rising, level high reg. (modal): falling, level	SF (Hari, 1977, p.64)
	d.	high reg. level [54/55] high reg. falling [52] low reg. level [13/113] low reg. falling [132]	SF (Hu, 1988; Hoshi, 2003)

Notes: s = short rhyme, l = long rhyme, reg. = register
SF = surface form UF = underlying form

Table B.1: Overview of mainstream LT (2/4) Tone system proposals

Qty		Tones	Remarks and references
5		High even 55 High falling 51 Low rising 13/14 Low rising 35 Light tone	SF (Sedláček, 1959, p. 245)
6	a.	high, low, high-high, high-falling, low-low, low-falling	SF (Chang and Chang, 1964)
	b.	(s) high level 5 (l) high level 55 (l) high falling 52/51 (s) low rising 23/2 (l) low rising 23/24 (l) low rising falling 231	SF/UF (L,H) (Dawson, 1980, p. 103), SF (Hu et al., 1982)
	c.	μ: H, L $\mu\mu$: HH, HL, LH, LML	SF(Gong, 2014)
8		H, LH, HS, LHH, HLS, LHS, HL, LHL	SF (Fang Hu, 2010)

Notes: s = short rhyme, l = long rhyme, reg. = register
SF = surface form UF = underlying form

Table B.2: Overview of mainstream LT (5/6/8) Tone system proposals

Qty		Tones	Remarks and references
0		HIGHF, lowf	Derived laryngeal feature (F/f) (Civera, 1970)
3	a.	L, M, H	Not view as pitch, but prominence (Hannah, 1912, p. 44)
	b.	Low resonant, Medial, High-pitched	Based on WT prefix and initials (Sandberg, 1894, p. 14)
	c.	Low, Medium, High	Based on WT prefix and initials (Bell, 1919, p. 18-19)
	d.	L, H, falling	(Goldstein et al., 2001)

Table B.3: Overview of non-mainstream LT Tone system proposals.

Appendix C

Surface realization of LT words

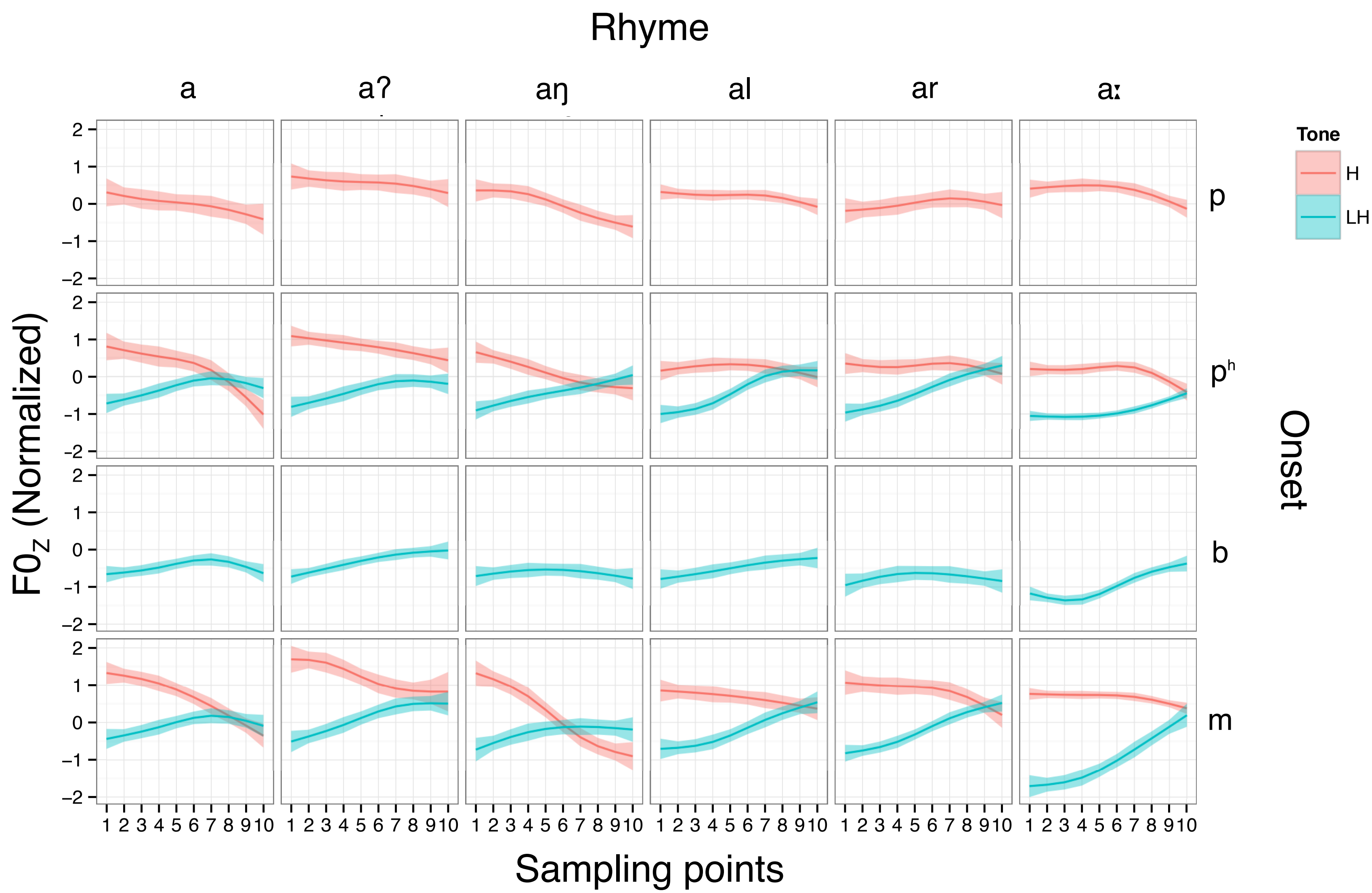


Figure C.1: Pitch pattern of monosyllable with various rhymes.

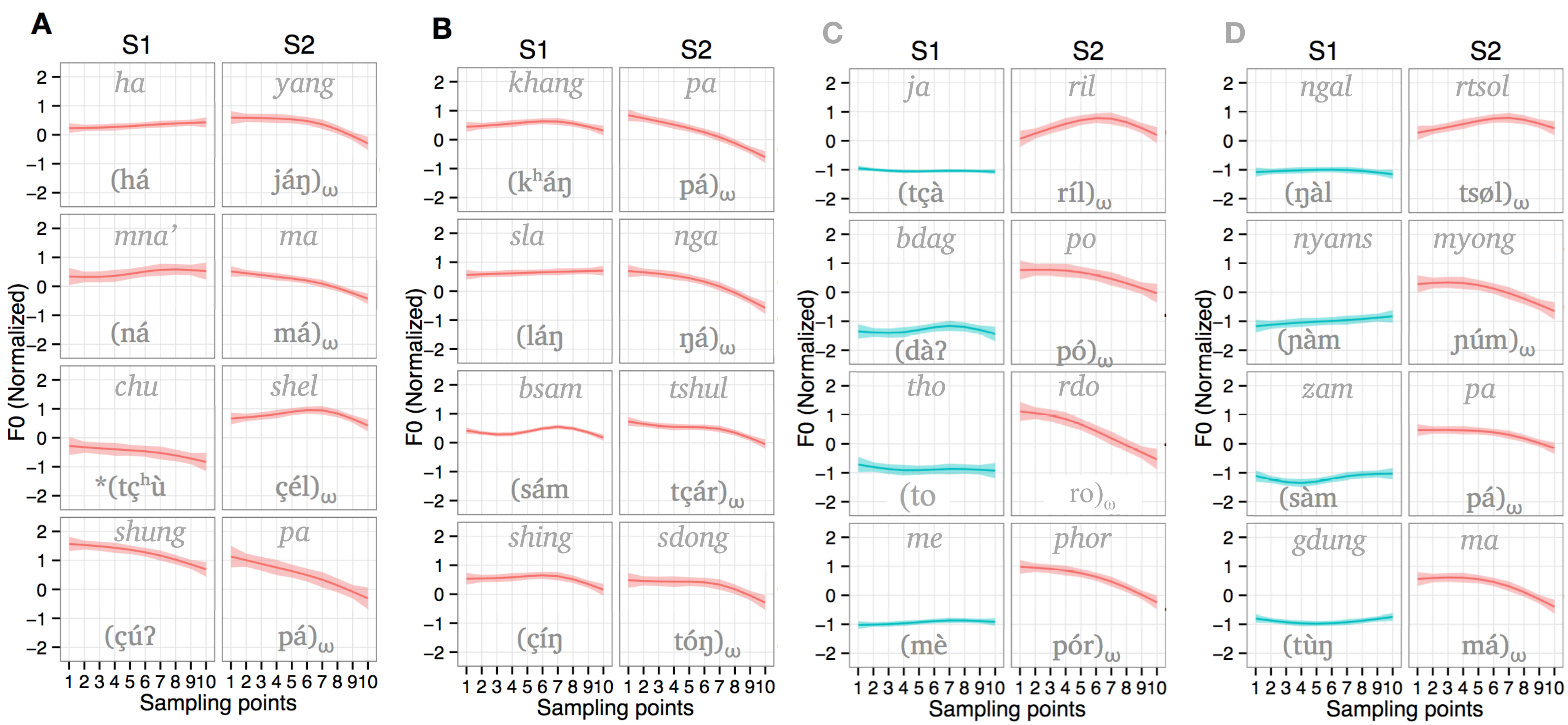


Figure C.2: Pitch pattern of LT disyllabic words.

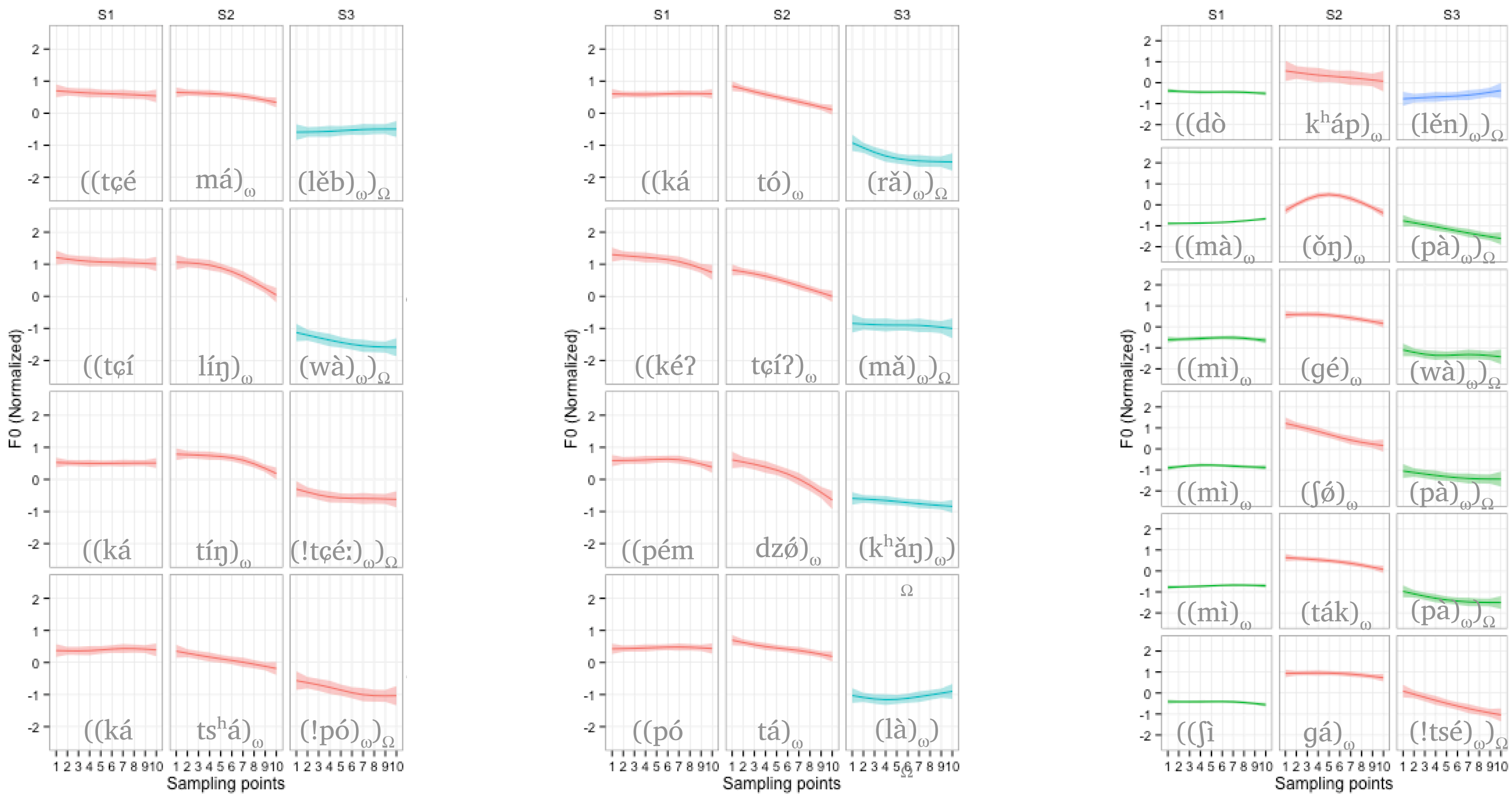


Figure C.3: Surface realization of LT trisyllabic words I.

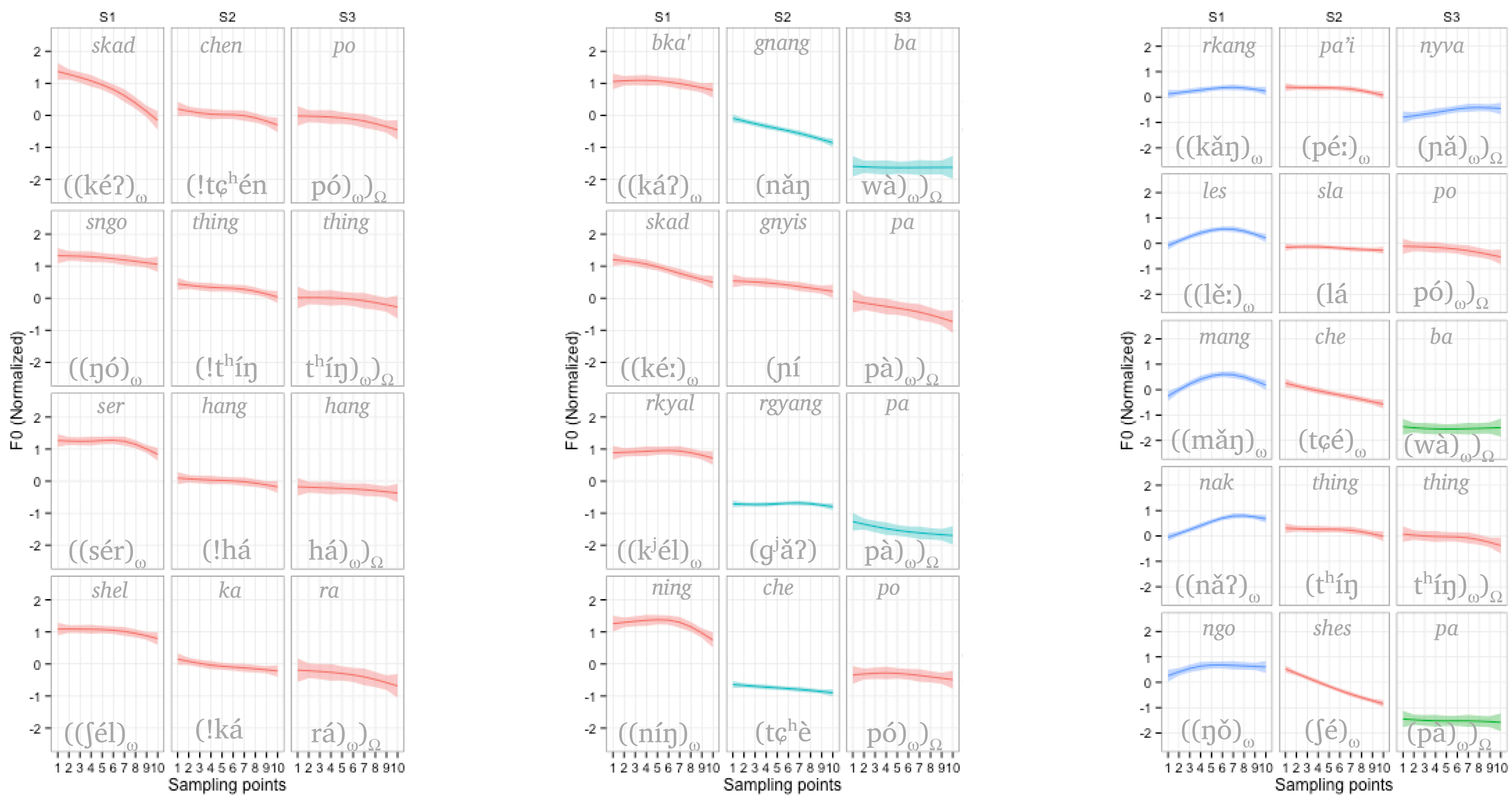


Figure C.4: Surface realization of LT trisyllabic words II.

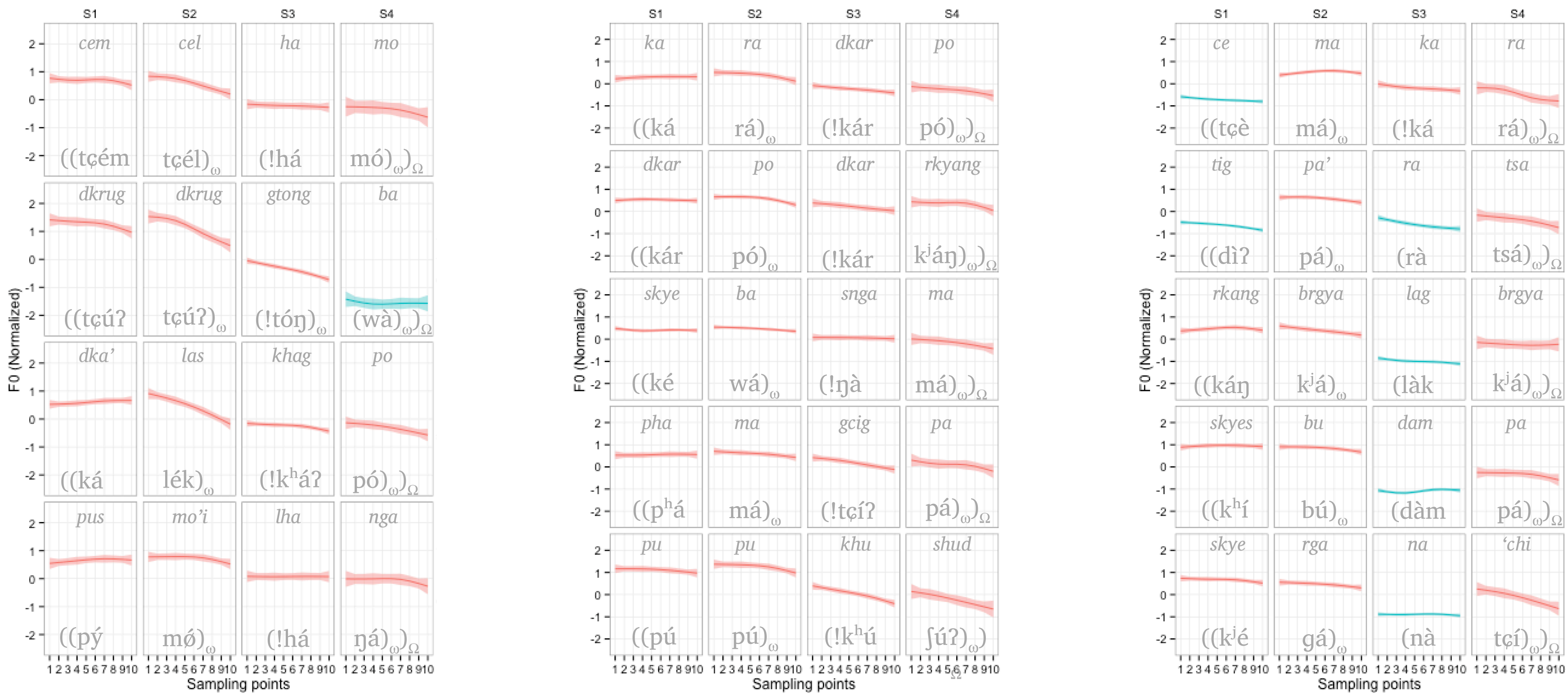


Figure C.5: Surface realization of LT quadrisyllabic words I.

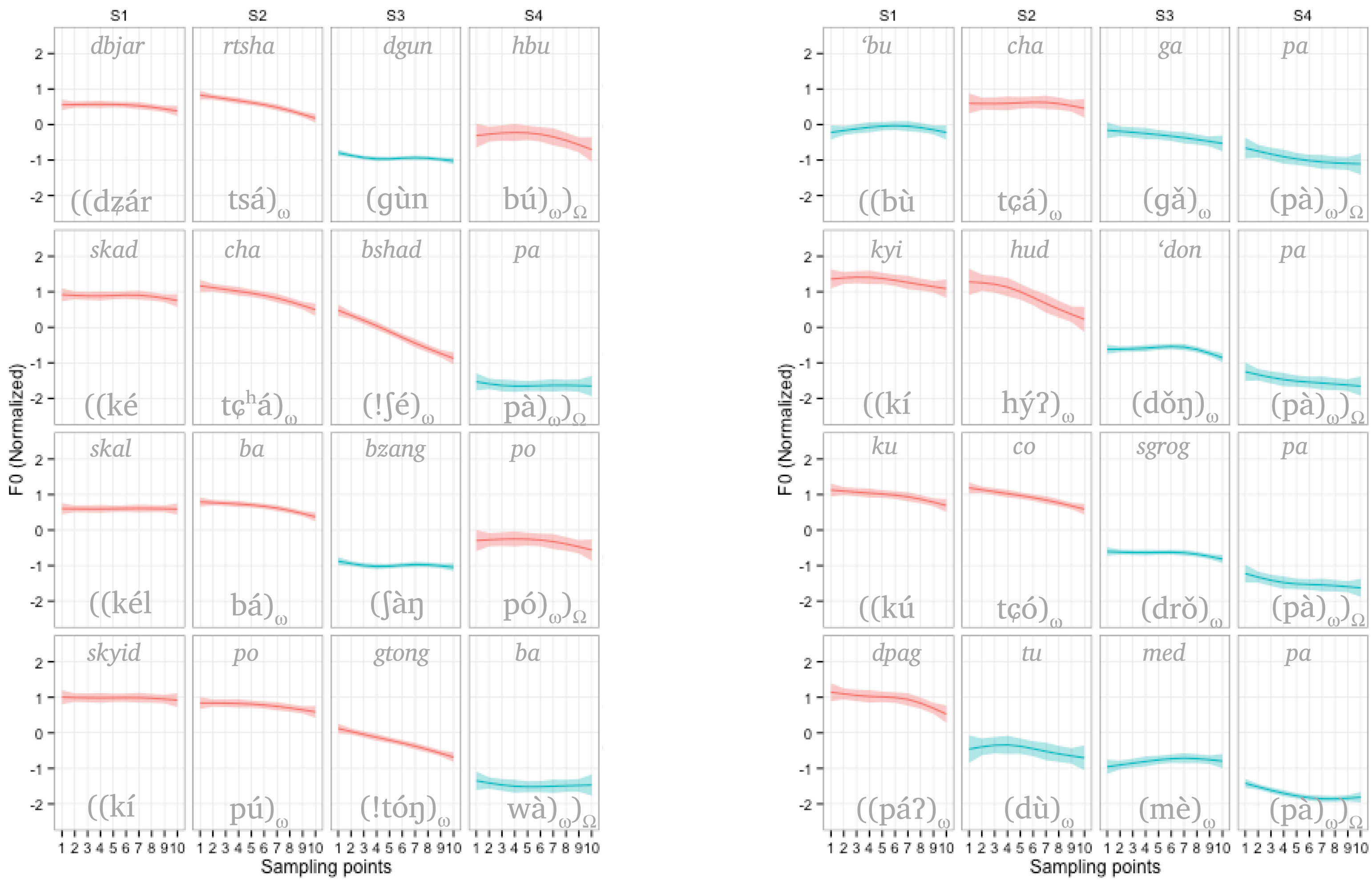


Figure C.6: Surface realization of LT quadrisyllabic words II.

Appendix D

Morphemes of LT polysyllabic words

WT	IPA	Morphemes	Gloss
$[((\sigma\sigma)_\omega(\sigma)_\omega)_\phi]$			
rdo khab len	[dò] [kʰáp] [lěn]	[stone] [needle] [pull/bring]	'magnet'
phjig ling ba	[tɕí.líŋ] [wà]	[foreign] [nominalizer]	'foreigner'
bka' drin can	[ká.tín] [tɕéː]	[thankfulness] [to have]	'the kind one'
katora	[ká.tó.rǎ]	[red copper basin *fr. Sanskrit*]	'red copper basin'
ka tsha po	[ká] [tsʰá] [pó]	[command] [hot] [adjectivizer]	'urgent'
skad cig ma	[kéʔ.tɕíʔ.mǎ]	[an instance *fr. Sanskrit*]	'a moment'
potala	[pó.tá.là]	[Pota] [?mountain pass]	'Potala palace'
gzhes ka rtse	[ʃì.ká.tsé]	[Shigatse]	'Shigatse'
phem ma leb	[tɕé] [má] [lěb]	[moth] [?] [at]	'moth'
$[((\sigma)_\omega(\sigma\sigma)_\omega)_\phi]$			
skad chen po	[kéʔ] [tɕʰén] [pó]	[voice] [very] [adjectivizer]	'loud, or noisy'
skad gynis pa	[kéː] [ɲí] [pà]	[voice/language] [two] [nominalizer]	'bilingual'
les sla po	[lěː] [lá] [pó]	[work?] [easy] [adjectivizer]	'easy'
nak thing thing	[nǎʔ] [tʰíŋ] [tʰíŋ]	[black] [penetrating?] [penetrating?]	'pitch black'
ning che po	[níŋ] [tɕʰé] [pó]	[compassion] [very] [adjectivizer]	'beautiful'
sngo thing thing	[ŋó] [tʰíŋ] [tʰíŋ]	[blue] [penetrating?] [penetrating?]	'blueish'
ser hang hang	[sér] [há] [há]	[yellow] [very?] [very?]	'yellowish'
$[((\sigma)_\omega(\sigma)_\omega(\sigma)_\omega)_\phi]$			
rkang pa'i nyva	[káŋ] [péː] [ɲǎ]	[leg] [nominalizer + GEN] [calf]	'calf'
rkyal rgyag pa	[kʲél] [gʲǎʔ] [pà]	[swim] [to do] [nominalizer]	'to swim'
ma 'ongs pa	[mà] [ǒŋ] [pà]	[NEG] [to come]	'future'
mi dge ba	[mì] [gé] [wà]	[NEG] [virtue] [nominalizer]	'non-virtue'
mi bzod pa	[mì] [ʃǿː] [pà]	[NEG] [bearable]	'unbearable'
mi rtag pa	[mì] [ták] [pà]	[NEG] [permanence]	'impermanence'
mang che ba	[mǎŋ] [tɕʰé] [wà]	[more] [very] [nominalizer]	'most'
ngo shes pa	[ŋǒ] [ʃé] [pà]	[face] [to know] [nominalizer]	'acquaintance'
shel kara	[ʃél] [kárá]	[rock] [sugar]	'rock sugar'

Figure D.1: Morphemes of LT Trisyllabic word stimuli.

WT	IPA	Morphemes	Gloss
$[((\sigma\sigma)_\omega(\sigma\sigma)_\omega)_\phi]$			
tig pa'ra tsa	[dìʔ.pá][rà][tsá]	[threatening][root of a horn]	'scorpian'
dkrug dkrug gtong ba	[tɕúʔ][tɕúʔ][tóŋ][wà]	[shake/stir][shake/stir][to do][nominalizer]	'stirring'
dka' las khag po	[ká.lék][kʰáʔ][pó]	[tired][difficult][adjectivizer]	'difficult'
dkar ra dkar po	[ká.rá][kár.pó]	[sugar][white][adjectivizer]	'white sugar'
rkang brgya lag brgya	[káŋ][kʲá][làk][kʲá]	[leg][hundred][hand][hundred]	'centipede'
dkar po dkar rkyang	[kár][pó][kár.kʲáŋ]	[white][adjectivizer]	'pure white'
skye ba snga ma	[ké.wá][ŋá.má]	[birth][previous]	'former life'
skyes bu dam pa	[kʰí.bú][dàm][pá]	[person/being][holy][nominalizer]	'saint'
skye rga na 'chi	[kʲé][gá][nà][tɕhí]	[birth][old age][sickness][death]	'life/cycle of life'
pha ma gcig pa	[pʰá][má][tɕíʔ][pá]	[father][mother][one][nominalizer]	'same parent'
pu pu khu shud	[pú.pú.kʰú.ʃýʔ]	[name of a Tibetan bird]	'name of a tibetan bird'
pus mo'i lha nga	[pý.mǿ][há.ŋá]	[knee][basin]	'knee cap'
cema kara	[tɕè.má][kárá]	[coarse][sugar]	'coarse sugar'
cem cel ha mo	[tɕém.tɕél][hámó]	[?][goddess/angel]	'butterflies'
$[((\sigma\sigma)_\omega(\sigma)_\omega(\sigma)_\omega)_\phi]$			
'bu cha ga pa	[bù][tɕʰá.gǎ][pà]	[small creature][grasshopper/insect][nominalizer]	'grasshopper'
dbjar rtsha dgun 'bu	[dʑár][tsá][gùn][bú]	[summer][plant][winter][worm]	'herbal medicine'
skal cha bshad pa	[ké.tɕʰá][ʃé][pà]	[converse][to speak][nominalizer]	'to talk'
skal ba bzang po	[kél.bá][ʃàŋ.pó]	[fortune][excellent][adjectivizer]	'excellent fortune'
skyid po gtong ba	[kí.pú][tóŋ][wà]	[happy/joy][to do][nominalizer]	'to be happy'
kyi hud 'don pa	[kʲí.hýʔ][drǒ/dǒn][pà]	[lamentation][to make][nominalizer]	'Oh!oh! wailing'
ku co sgrog pa	[kú.tɕó][drǒ/dǒn][pà]	[noise][to make][nominalizer]	'to make noise'
$[((\sigma)_\omega(\sigma)_\omega(\sigma)_\omega(\sigma)_\omega)_\phi]$			
dpag tu med pa	[páʔ][dù][mè][pá]	[fathomable][connector][NEG][nominalizer]	'Unfathomable'

Figure D.2: Morphemes and prosodic phrasing of LT Quadrisyllabic word stimuli.

Appendix E

Phonological derivation of LT words

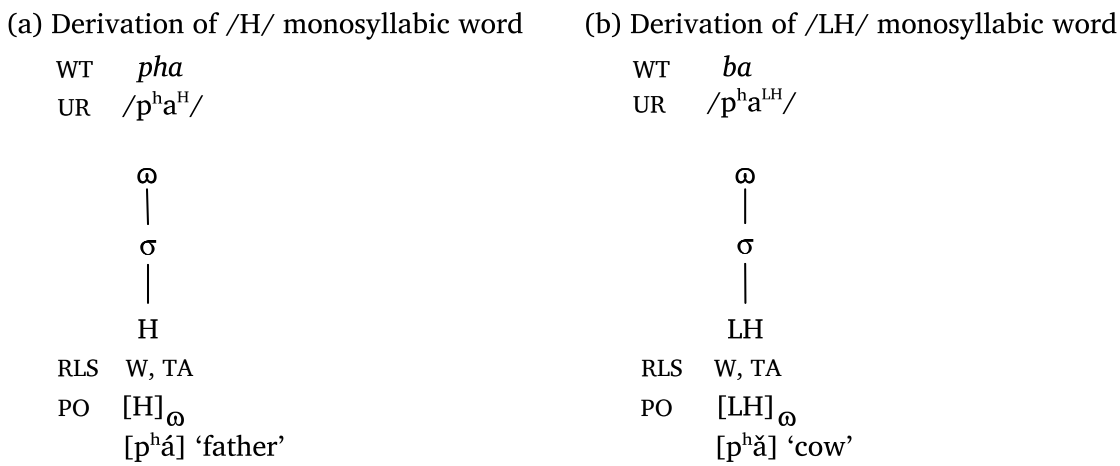


Figure E.1: Derivation of monosyllabic words.

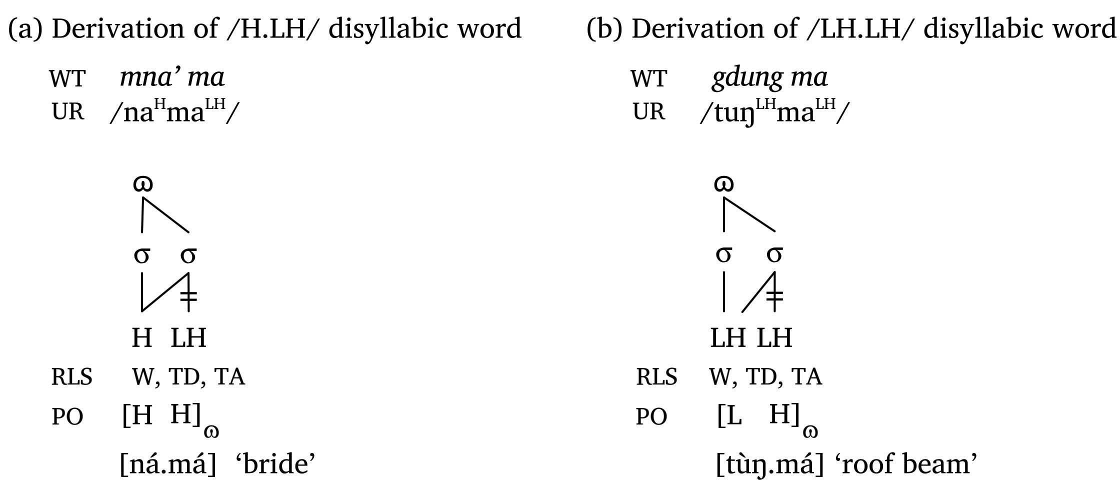


Figure E.2: The derivation of disyllabic words.

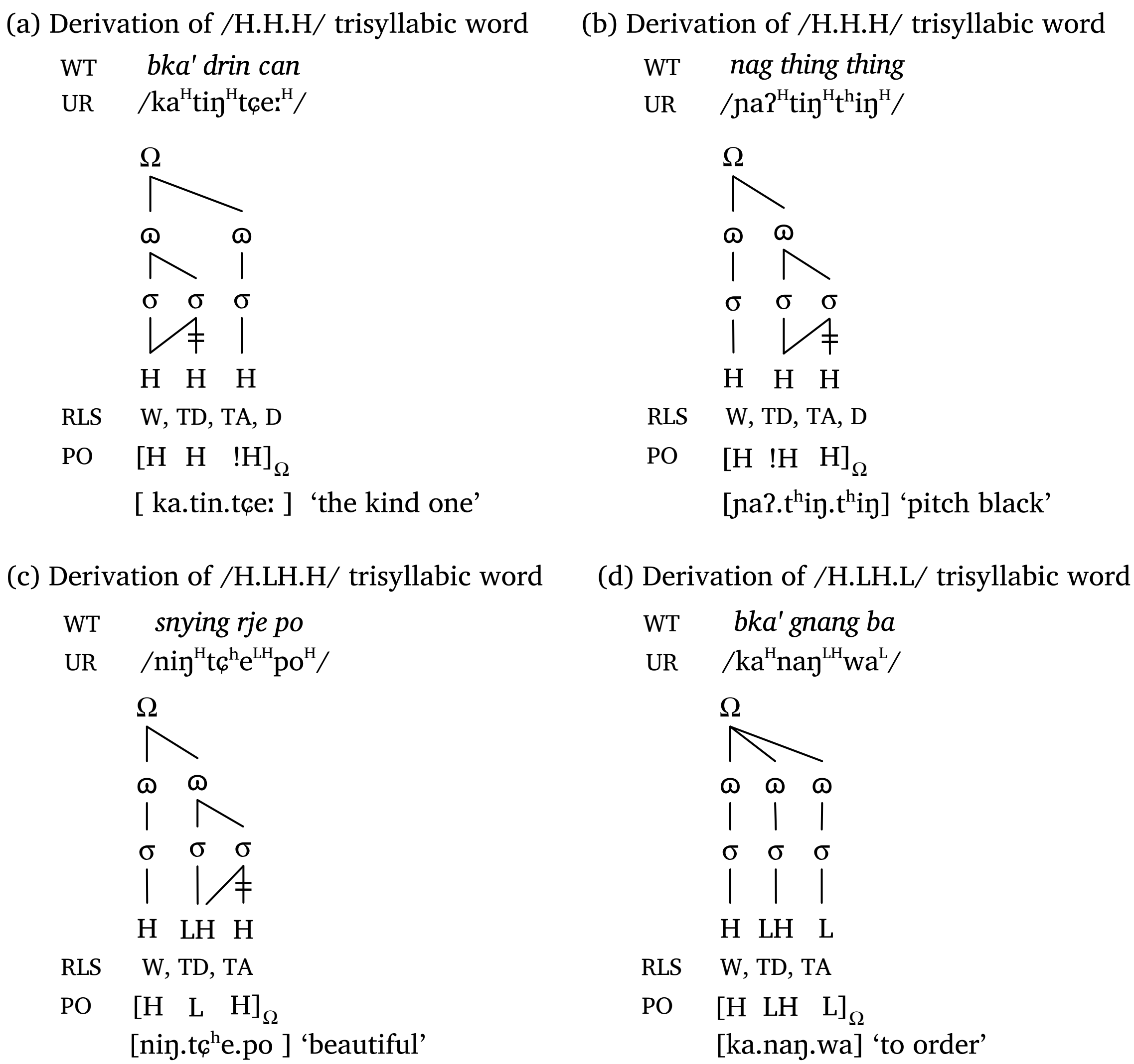


Figure E.3: LT trisyllabic words with word-initial H.

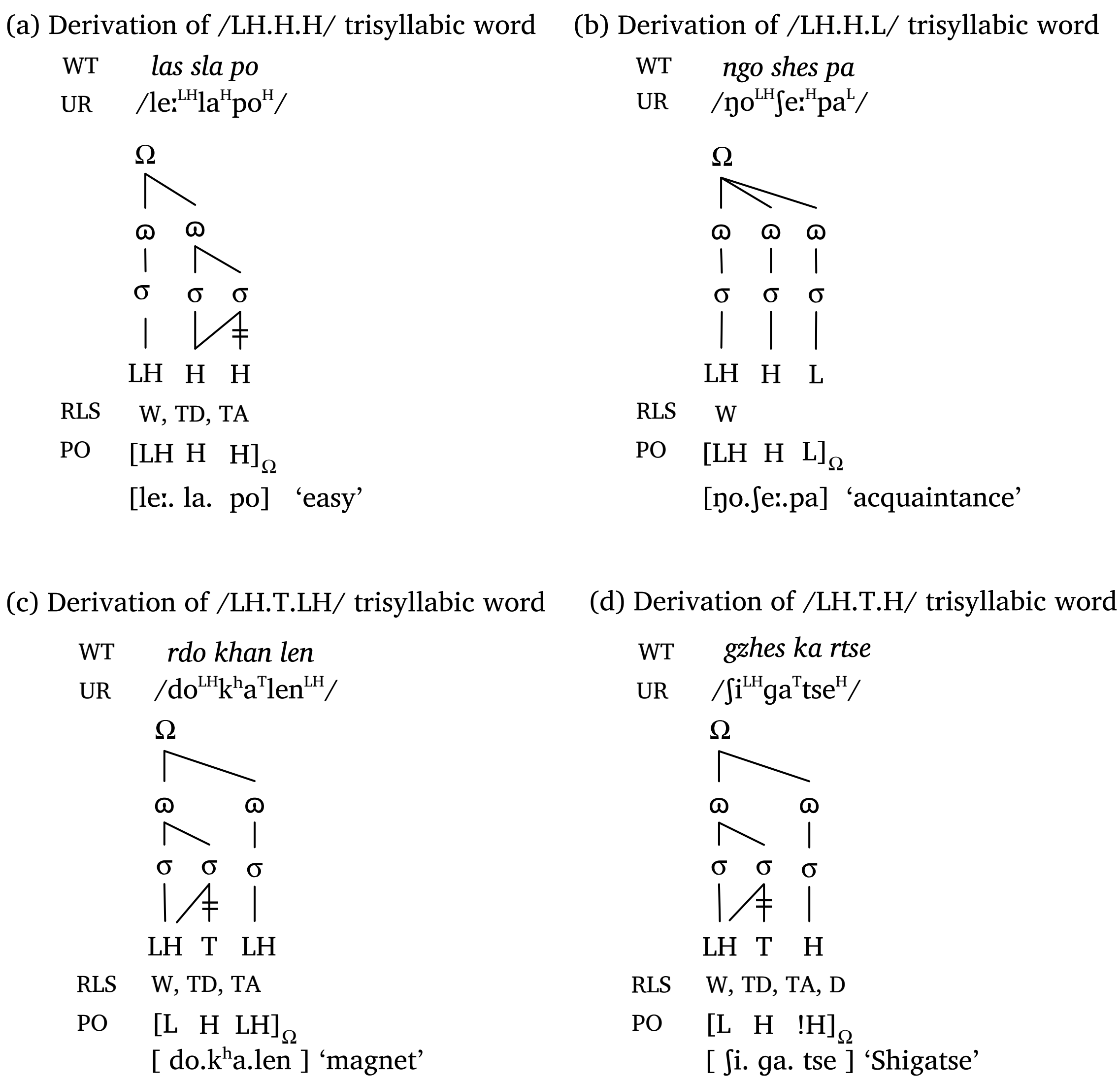


Figure E.4: LT trisyllabic word examples with word-initial H.

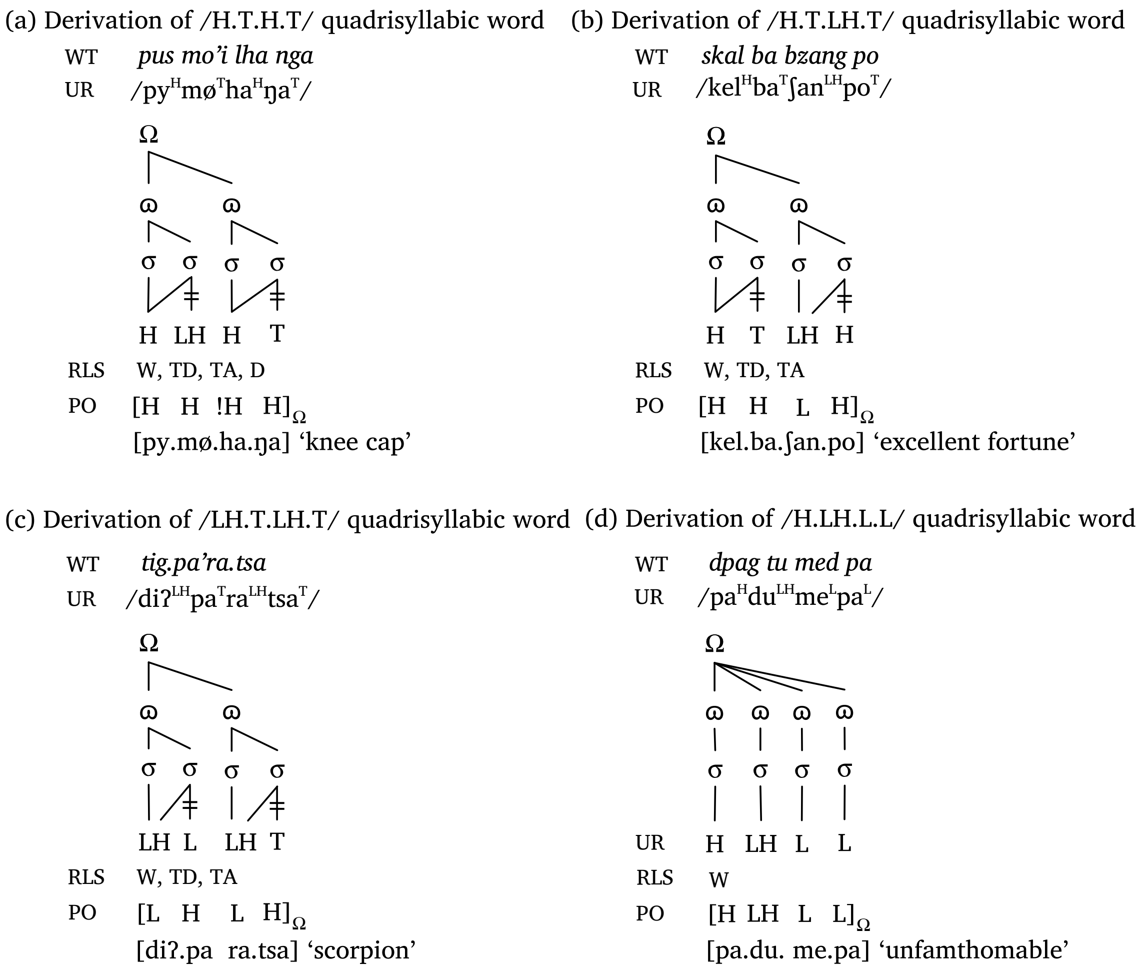


Figure E.5: LT quadrisyllabic word examples.