

A phonological model of Uyghur intonation

Travis Major and Connor Mayer

Abstract

This chapter presents an intonational model for Uyghur (Turkic: China). First, it is demonstrated on the basis of acoustic measurements that Uyghur is a stress language that only uses edge-marking intonation. Although such languages are attested in the literature, this is the first autosegmental-metrical model of such a language. Next, a phonological model of Uyghur intonation is provided in the autosegmental-metrical framework, where the basic prosodic constituents and edge marking tones are described. We finish by applying this model to various sentence types and outlining areas for future research.

Keywords: Uyghur, intonation, prosody, autosegmental-metrical, stress

6.1 Introduction

Uyghur (ISO 639-3: *uig*) is a southeastern Turkic language with roughly ten million speakers in the Xinjiang Uyghur Autonomous Region in the People’s Republic of China, and neighboring regions such as Kazakhstan, Uzbekistan, and Kyrgyzstan. It is a synthetic, agglutinating language with SOV word order and a rich case marking and agreement system. It is typologically most similar to modern Uzbek (Engesæth, Yakup, & Dwyer, 2009/2010).

The goal of this chapter is to provide a model of the intonational phonology of Uyghur in the autosegmental-metrical (AM) framework (e.g. Pierrehumbert, 1980; Beckman & Pierrehumbert, 1986; Ladd, 1996/2008), extending previous work (Major & Mayer, 2018). The AM theory proposes that the continuous pitch contour of utterances can be broken down into a string of discrete pitch targets that consist only of high (H) or low (L) tones, or complex combinations of the two (e.g. LH or HL). These tones are associated with particular parts of the segmental string in two ways: (i) head-marking tones, or pitch accents, associate with a prominent syllable or mora; and (ii) edge-marking or boundary tones associate with the edges of prosodic constituents. Phonetic interpolation determines the pitch contour between tonal targets. The AM model provides a useful set of theoretical assumptions for analyzing the intonational systems of languages, and the analysis of Uyghur in this framework will allow for typological comparisons with the systems of other languages.

This chapter is structured as follows. We first describe the relationship between stress and pitch in Uyghur: we find evidence supporting past descriptions of Uyghur as a stress language with only edge-marking intonation. This makes Uyghur unique among languages that have been modeled in the AM framework. We then outline the basic prosodic constituents and edge marking tones of our proposed model. We end by discussing the intonational properties of a range of different sentence types, including more naturalistic speech.

6.2 Background

There has been little work to date on intonation in Uyghur, although there has been some on the prosodic systems of related languages like Turkish and Chuvash.

6.2.1 Past work on Turkic prosody

The status of lexical stress in Turkish has been heavily debated in the literature. Turkish has been traditionally analyzed as a stress language (e.g., Lees, 1961; Kaisse, 1985; Barker, 1989; Inkelas, 1999; Inkelas & Orgun, 1998; Kabak & Vogel, 2001; Ipek & Jun, 2013; Ipek, 2015), while others have argued that Turkish is a lexical pitch accent language (e.g., Levi, 2005; Kamali, 2011; Günes, 2015). It has been noted that the nuclear pitch accent in Turkish is realized in a more compressed pitch range than the pre-nuclear pitch accent (Kamali, 2011; Kan, 2009). More recently, Ipek and Jun (2013) show that the nuclear pitch accented word is marked on its left edge by an H tone in addition to the pitch range compression, while there is an additional H target associated with the right edge of NPs and PPs.

The study of the Turkic language Chuvash in Dobrovolsky (1999) also suggests that it is a stress language, with duration and intensity serving as important cues. No correlative measures of pitch were done, however.

6.2.2 Past work on Uyghur stress

The status of stress in Uyghur is not well understood. Hahn (1991a, 1991b) claims that stress is assigned predictably based on a number of prosodic factors, and is reflected by increases in pitch, duration, and intensity. Under this account, the final heavy syllable (CVV, CVC, CVVC, etc.) that occurs in ultimate or penultimate position will receive stress. If no such heavy syllable is present, stress defaults to the final syllable of the word. Heavy syllables that occur before the penultimate syllable receive secondary stress. Hahn also notes the existence of suffixes that attract or repel stress to the preceding syllable (so-called “pre-stressing suffixes”).

Engesæth et al. (2009/2010) agree that Uyghur is a stress language that defaults to word-final stress, but suggest that the interaction of stress with syllable weight is a tendency, and not a rule. They suggest that stress tends to fall on the first heavy syllable in a word (e.g., *tápshuruq* ‘homework’, *murékkep* ‘complicated’), but note exceptions to this (e.g., *Turpán* ‘Turpan’). In addition, they suggest that loan words maintain stress patterns from the original language (e.g., *gimnástika* ‘gymnastics’ from Russian; *bála* ‘disaster’ from Farsi; cf. the native Uyghur word *balá* ‘child’). They also suggest that the primary acoustic correlate of stress is duration, not F0 or intensity.

A series of production and perception experiments done by Yakup (2013) support this claim. These experiments targeted stress minimal pairs or near minimal pairs such as *bála* ‘disaster’ and *balá* ‘child’ in both single word utterances and continuous speech, and showed that only duration served as a significant correlate of stress location. F0 and intensity were not significantly correlated with perceived prominence by speakers, which suggests that Uyghur uses a more limited set of acoustic features to mark stress than other stress languages. However, Yakup also found that speakers frequently disagreed as to which syllables were stressed in many words, indicating that stress may not be robustly perceived or produced, even by native speakers.

Özçelik (2015), on the other hand, presents formal and experimental evidence that Uyghur is a predominantly footless language that features intonational prominence on the right edge of prosodic words. He is careful to state that this prominence is not stress, but a boundary tone at the right edge of the prosodic word, and shows no accompanying increase in duration. Suffixes that generate prominence on non-final syllables (i.e., those that either attract or repel stress) are cases of true stress: they are claimed to have underlying trochaic foot structure, which necessitates footing the word in such a way that produces non-final stress. These exceptionally stressed syllables are claimed to have greater F0 and duration. Although some of the claims from this paper are corroborated by previous research and the current paper (e.g., intonational prominence on the rightmost edges of words, the presence of idiosyncratic stress), the study contains a large number of methodological issues that undermine the validity of many of its claims: to name just two, the claim that exceptional stress is correlated with increased F0 is based on differences in production between declarative and interrogative sentences, with no attempt to control for results arising from sentence or utterance level intonational properties; and the claim that final syllables are not lengthened is based on a statistical analysis of measurements made from a single token of a single word (*paqa*, ‘frog’) from just five speakers.

Though we acknowledge there is much work to be done to better understand stress in Uyghur, we adopt an account that is broadly consistent with Engesæth et al. (2009/2010) and Yakup (2013): the only reliable acoustic correlate of stress in Uyghur is increased duration, and although stress prefers to fall at the right edge of a word and on heavy syllables, there are many exceptions, particularly in loan words. This means that stressed syllables in Uyghur cannot be identified from the pitch contour of an utterance. This differs from Turkish, where intonational tones do associate with stressed syllables.¹ In the AM theory of intonation, intonational tones mark lexical heads (i.e. stressed syllables) and the edges of prosodic units. That is, if a language has stress, the stressed syllable is expected to be marked by intonation. Therefore Uyghur would be somewhat unusual from the perspective of the prosodic typology outlined in Jun (2005), which does not identify any languages that have stress word prosody but only edge marking intonation.²

Such languages are not completely unknown in the literature, however: Lindström and Remijsen (2005) suggest that Kuot, a non-Austronesian language of Papua New Guinea, displays similar properties, with strong effects of duration for word stress and F0 for intonational marking, but no interaction between the stress and intonational systems. Similarly, Kisseberth and Abasheikh (2011) report that Chimwiini intonation is independent of vowel length, which correlates with stress. The description of the Turkic language Chuvash in Dobrovolsky (1999) also has some intriguing suggestions of mismatches between stress and intonation, but there is not sufficient data presented to determine whether pitch accents are present or not.

If it is indeed the case that Uyghur is a stress language with only edge marking intonation,

¹ Stress in Turkic languages was historically realized on the root, but eventually got shifted to the final syllable with certain exceptions. This tendency for final stress is robust across Turkic, but each language has developed a unique system (Menges, 1995).

² An additional possibility that bears mention here is that Uyghur has no stress at all, but simply a distinction between long and short vowels (e.g., Hahn, 1991a, 1991b, though Hahn also assumes stress). It is unclear how to differentiate empirically between phonemic contrasts in vowel length and lexical stress whose sole acoustic correlate is duration. We assume the latter for several reasons: no Uyghur orthography distinguishes between long and short vowels, speakers often disagree as to whether a particular syllable is prominent or not, and minimal pairs that differ in stress position/vowel length appear to be relatively uncommon compared to other languages with a recognized phonemic length distinction.

formalizing this into an AM model will be a useful step towards expanding our typological inventory of intonational systems. The next section will provide experimental evidence that this is the case, which will subsequently be used to motivate our proposed model of Uyghur intonation.

6.3 Stress and intonation in Uyghur

Yakup (2013) shows that stress in Uyghur is reflected only by vowel length, not pitch or intensity. This suggests that although Uyghur can be described as a stress language, stress and intonation are independent. In this section, we use acoustic measurements to show that vowel duration, but not F0, is the only acoustic correlate of stress, while F0 is used to mark the boundaries of prosodic constituents (see Section 6.4). This is largely a replication of Yakup (2013), but we introduce one important addition: we examine the same words in both sentence-initial and sentence-medial positions, while the studies in Yakup (2013) only looked at words in isolation and in sentence-medial contexts. We will show later that there is a tendency for different prosodic constituents to be used for words in these positions, and that this difference is reflected in the acoustic measurements.

6.3.1 Data collection

Data for the acoustic stress study were collected from eight adult speakers of Uyghur, four male and four female. Four of the speakers (2M, 2F) are from Xinjiang: three from the greater Urumqi area, and one from Qashqar. The other four (2M, 2F) are from the Almaty region in Kazakhstan. All speakers were educated in Uyghur and raised speaking primarily in Uyghur. The speakers from Xinjiang are all currently pursuing post-secondary degrees in the United States or working as academics, while the speakers in Kazakhstan were teachers at a Uyghur high school in Almaty.

Sentences were elicited by having the consultants read from a randomized list prepared by the authors. Sentences were checked for grammatical acceptability with consultants before recording them. The recordings for speakers from Xinjiang were made in sound booths in the UCLA and University of Kansas departments of linguistics. The recordings for speakers from Almaty were made in a quiet room at the Uyghur high school.

6.3.2 Stimuli

We tested the independence of stress and intonation in Uyghur by eliciting a series of minimal and near-minimal stress pairs from Yakup (2013) for which speakers showed a high level of agreement about stress location. These pairs were elicited in both sentence-initial and sentence-medial position in the following carrier phrases:

- *X bek yaxshi söz* – “*X* is a good word”
- *Mahinur X deydu* – “Mahinur will say *X*”

These carrier phrases were chosen to accommodate the various parts of speech of the target words. Although these carrier phrases set off the target words somewhat from the rest of the sentence, this applies equally to all target words, and hence comparison between them is justified. Our target words are shown in Table 6.1. Stress is indicated by capitalization. This design allowed

a balanced number of stressed and unstressed vowels in similar contexts and with mostly the same vowel quality. This resulted in 16 tokens per speaker, for a total of 128.

Figs. 6.1 and 6.2 contrast *Acha* “elder sister” and *aCHA* “branching” in sentence-initial position. Compare the relative duration of the two vowels in each word: the first and second vowels in Fig. 6.1 are 131 ms and 78 ms respectively, while in Fig. 6.2 they are 80 ms and 118 ms.³

Word 1	Gloss 1	Word 2	Gloss 2
DAka	gauze	daLA	plain
BAza	base	baHA	price
DAcha	villa	daDA	father
DOra	medicine	doQA	forehead
CHAsa	square	chaTAQ	problem
Acha	elder sister	aCHA	branching
BAla	child	baLA	disaster
Ara	fork	aRA	between

Table 6.1: Near-minimal and minimal stress pair target words.

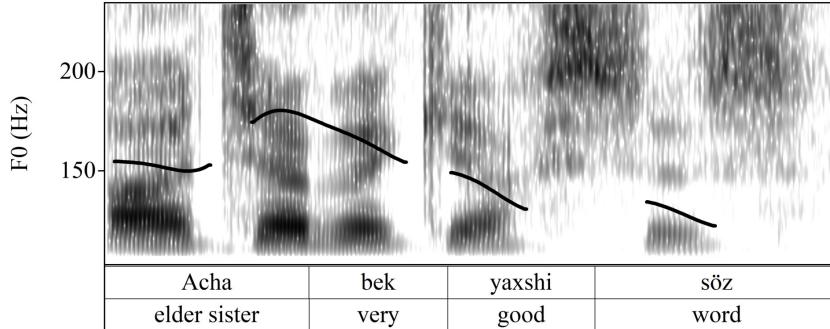


Figure 6.1: Pitch track of word-initial stress in sentence-initial position.

6.3.3 Analysis

The two vowels in each word were segmented using Praat (Boersma & Weenink, n.d.), and the average intensity (in dB), average F0 (in Hz), and duration (in seconds) were extracted. We ran three linear mixed effects models using the *lme4* package in R (R Core Team, 2017; Bates, Mächler, Bolker, & Walker, 2015). Significance values were generated using the *lmerTest* library (Kuznetsova, Brockhoff, & Christensen, 2017). Our dependent variables were duration, intensity, and pitch respectively. Our independent variables were stress (stressed or unstressed), position in the word

³Note that the target word in Figs. 6.1 and 6.2 is focused, and the material following the target word is deaccented, resulting in a fairly linear decline in pitch. We return to this phenomenon in the discussion section.

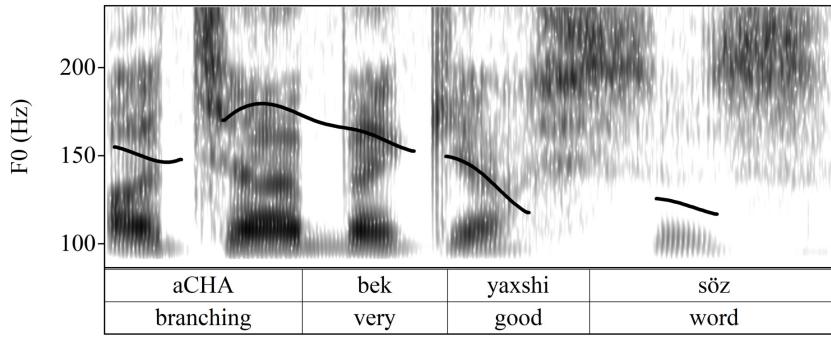


Figure 6.2: Pitch track of word-final stress in sentence-initial position.

(first or second syllable), and position of the word in the sentence (initial or medial). We included random intercepts for word and participant and random slopes for word and participant over stress. Additional random intercepts for vowel identity and speaker background (Xinjiang vs. Almaty) did not improve the fit of the models.

6.3.4 Results

The duration model showed a significant main effect of stress ($\beta = -0.018, t = -3.042, p < 0.01$), with unstressed vowels being significantly shorter, a significant main effect of position in the word ($\beta = 0.012, t = 2.472, p < 0.05$), with vowels in the second syllable of words being significantly longer, and a significant main effect of position in sentence ($\beta = -0.008, t = -2.29, p < 0.05$), with vowels in sentence-medial words being significantly shorter. Duration values are plotted in Fig. 6.3.

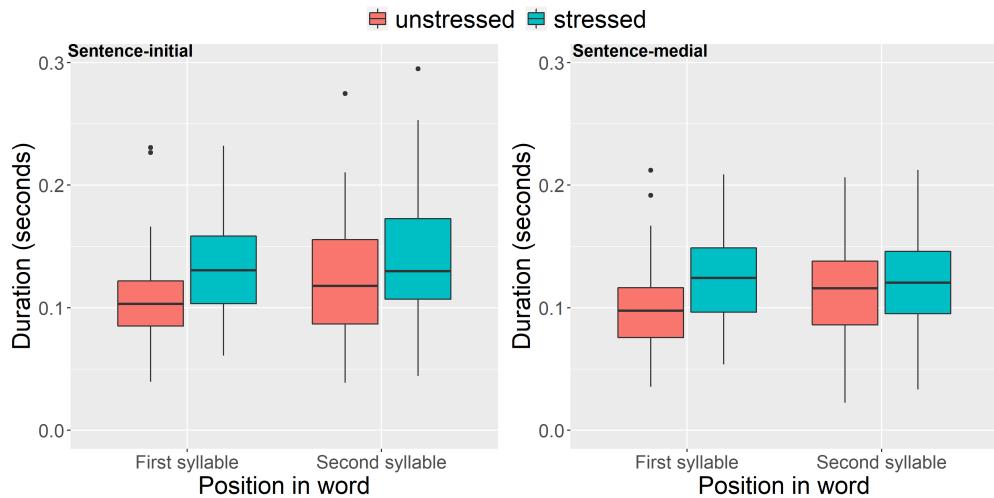


Figure 6.3: Vowel duration broken down by position of word in sentence, position of vowel in word, and whether or not the vowel is stressed. Values are reported in seconds.

The only significant effect in the intensity model was a main effect of position in sentence ($\beta = -1.194, t = -3.058, p < 0.01$), with words in sentence-medial position having significantly lower intensity. There was no significant effect of stress on intensity. Intensity values are plotted in Fig. 6.4.

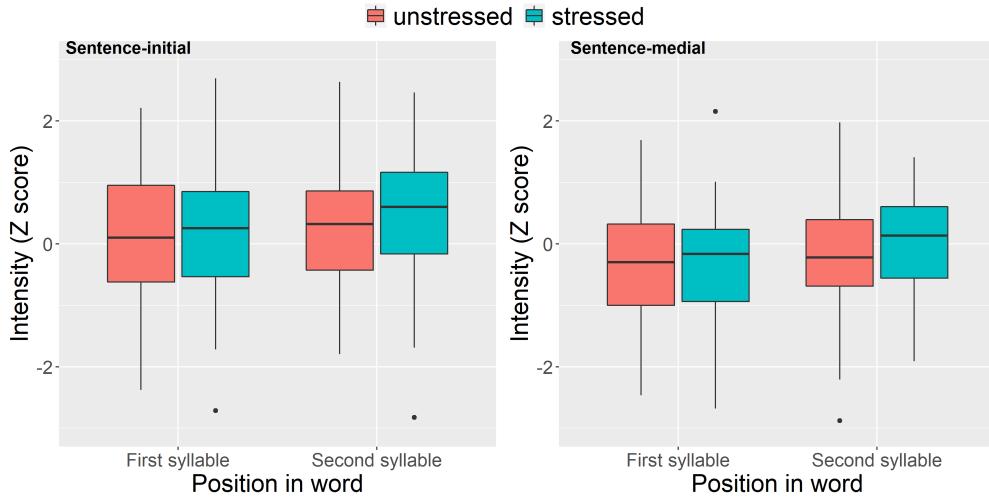


Figure 6.4: Average vowel intensity broken down by position of word in sentence, position of vowel in word, and whether or not the vowel is stressed. Values are reported in Z scores rather than dB to facilitate comparison between speakers.

The pitch model showed a significant main effect of position in the word ($\beta = 40.021, t = 10.577, p < 0.001$), with word-final syllables having a significantly higher pitch, and a significant main effect of position of the word in the sentence ($\beta = -10.737, t = -3.450, p < 0.001$), with vowels in sentence-medial words having a lower pitch, suggesting declination. In addition, there was a significant interaction between position in word and position in sentence, with word-final vowels in sentence-medial words having a significantly lower pitch ($\beta = -15.327, t = -3.352, p < 0.001$). There were no significant effects of stress on pitch.

These results show that stress location is a significant predictor of duration, but not pitch. Pitch, rather, is predicted by the position in the word (word-final syllable > word-initial syllable) and the position of the word in a sentence (sentence-initial > sentence-medial), reflecting the edge-marking function of pitch. These results support treating Uyghur as a stress language with only edge-marking intonation, which serves as the basis for the model described below. In addition, the interaction seen in the pitch model between position in word and position in sentence indicates that word-final syllables in sentence-medial position have lower pitch than predicted by declination. This will also be important for the model presented below.

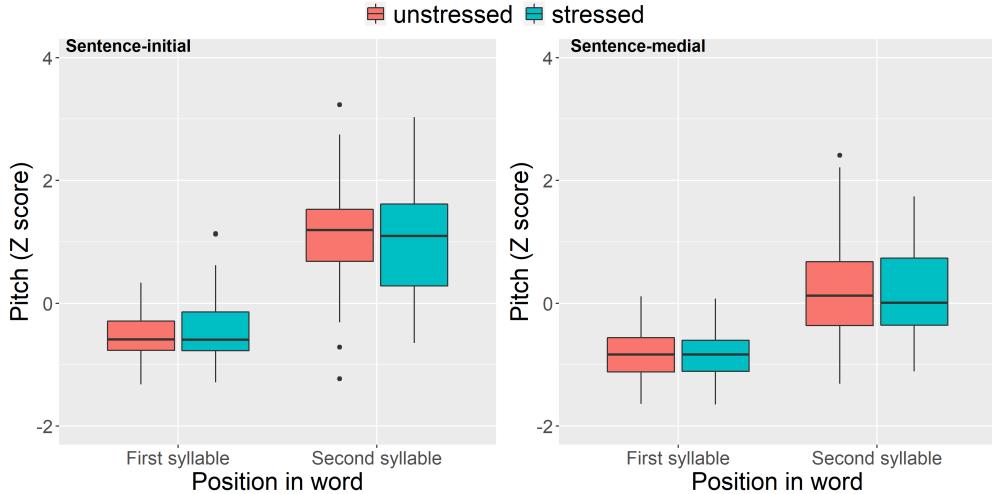


Figure 6.5: Average vowel pitch broken down by position of word in sentence, position of vowel in word, and whether or not the vowel is stressed. Values are reported in Z scores rather than Hz to facilitate comparison between speakers.

6.4 The Intonational Phonology of Uyghur

6.4.1 Data collection

The data that serves as the basis for the model presented in the remainder of the paper was collected from six native Uyghur speakers from Xinjiang (3M, 3F). Elicited sentences were constructed to contain as many sonorants and voiced sounds as possible to allow extraction of a clear F0 contour. Sentences were always elicited using a preceding question to provide an appropriate context. Broadly speaking, these preceding questions fall into three groups that differ in the kinds of focus they are intended to produce.

- **Neutral declarative example:**

<i>Videoda néme boldi?</i>	'What happened in the video?'
<i>Meryem Alimgha méwe berdi</i>	'Meryem gave fruit to Alim'

- **Wh-question example:**

<i>Kim Alimgha méwe berdi?</i>	'Who gave fruit to Alim?'
<i>Meryem Alimgha méwe berdi</i>	'Meryem gave fruit to Alim'

- **Contrastive focus example:**

<i>Mahire Alimgha méwe berdimu?</i>	'Did Mahire give fruit to Alim?'
<i>Yaq, Meryem Alimgha méwe berdi</i>	'No, <u>Meryem</u> gave fruit to Alim'

Not all data points were elicited from all participants, but the model presented here is consistent with the observed data from all participants. In cases where elicitation was done with a single participant, the investigator read the preceding questions. When elicitation was done with more than one Uyghur speaker, speakers took turns reading question and answer pairs. We also include several more naturalistic utterances towards the end of the paper.

6.4.2 Prosodic levels

Based on evidence from the distribution of intonational tonal targets, as well as phonological and syntactic properties, we argue that the Uyghur intonational system is sensitive to four distinct levels of prosodic constituency: the prosodic word (Pwd) the accentual phrase (AP), the intermediate phrase (ip), and the intonational phrase (IP). A schematized representation of these constituents is shown in Fig. 6.6.

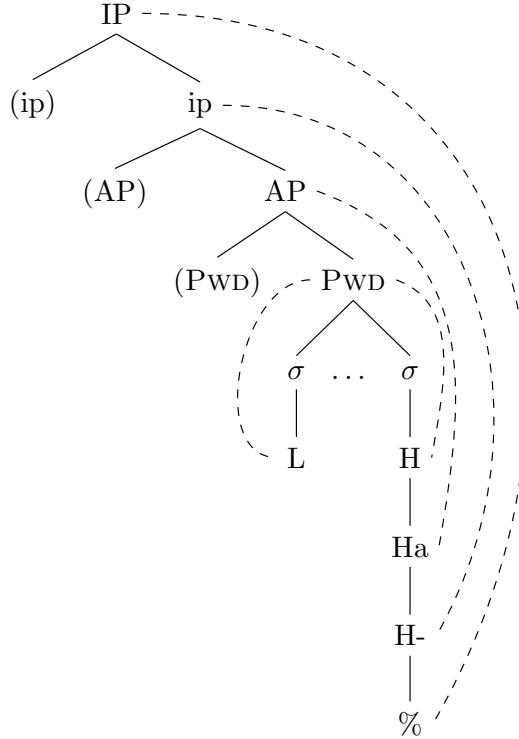


Figure 6.6: A schematic representation of the proposed prosodic hierarchy for Uyghur. Links between prosodic constituents and boundary tones are indicated by dotted lines. Prosodic tones associated with higher prosodic constituents overwrite tones associated with lower ones (i.e. % >> H- >> Ha >> H)

The following sections will introduce prosodic levels in pairs for expositional purposes, since it is informative to focus on the differences between lower and higher levels.

6.4.3 The prosodic word and the accentual phrase

The prosodic word is the lowest prosodic level above the syllable. Prosodic words generally consist of a single stem and its affixes (though see Section 6.6 for some examples of larger prosodic words), and are invariably marked on the left edge by an L tone. There are no clear cases in our data of a word beginning with anything but an L boundary tone, though in short words or in the presence of reduced or elided vowels in rapid speech this tone may manifest as a pitch plateau rather than a decrease when between two high tones (see, e.g., *bilen* in Fig. 6.14). There is also no clear evidence that higher prosodic levels have associated left edge-marking tones that override this tone. This is therefore the only left edge-marking tone in our model. An H boundary tone is

associated with the right edge of prosodic words. This tone is only observable when two prosodic words are adjacent to one another and constitute a single accentual phrase and, like the L tone, may manifest as a pitch plateau between two low targets on short words or in rapid speech (see, e.g., *mangidu* in Fig. 6.14).

The accentual phrase (AP) is the next highest level of prosodic constituency. Like prosodic words, APs are marked on the right edge by a high tone, which we write as Ha. This high tone differs from the H tones associated with prosodic words in two respects: first, Ha boundary tones have a higher F0 peak than H boundary tones. Second, Ha tones always produce a marked rise rather than pitch plateau.

Our data suggest that APs are generally no longer than two prosodic words: that is, we generally see an rhythmic alternation between lower and higher pitch peaks at the ends of words in an utterance. Because boundary tones associated with higher constituents overwrite those associated with lower ones (see Fig. 6.6), the right edge of the final prosodic word of an AP will always bear tonal marking associated with an AP, or a higher constituent.

Fig. 6.7 contains two APs that contain multiple prosodic words: *nérwa momay* “nervous grandmother” and *ram onglidi* “fixed a frame” (we will discuss the Ha/H- annotation in Section 6.4.4). The first word in each of these APs exhibits an L tone on the left edge and an H tone on the right edge. Notice that the H tone on the right edge of the adjective *nérwa* “nervous” is somewhat lower than the Ha tone on the right edge of *momay* “grandmother”: this difference is more striking when expected pitch declination is considered. There is a small pitch rise at the end of *ram* “frame” corresponding to the H boundary tone, followed by an F0 descent across the verb resulting from the L% tone associated with declarative utterances (see Section 6.4.6).

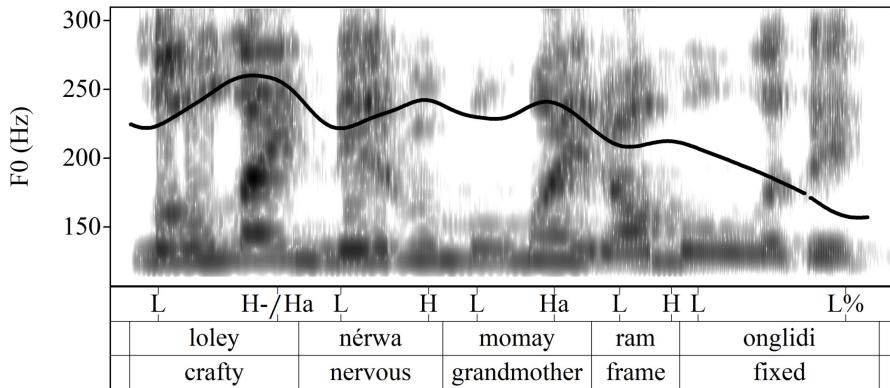


Figure 6.7: Pitch track, spectrogram, and annotation for the sentence *loley nérwa momay ram onglidi* “the crafty, nervous grandmother fixed the frame.”

In addition to the tonal patterns described above, there are segmental phonological processes that occur at prosodic word boundaries within an AP, but not at AP boundaries. One of these processes is *vowel reduction*: although there are many lexical and morphological complications that must be considered (see, e.g., Mayer, 2020), broadly speaking this process requires that the vowels *a* (IPA: /ɑ/) and *e* (IPA: /æ/) raise to *i* (IPA: [ɪ]) in medial, open syllables.

bala	‘child’	bali-ni	‘child-ACC’
apa	‘mom’	api-si	‘mom-3.SG.POS’
angla-sh	‘listen-GER’	angli-di	‘listen-3.SG.PAST’
qara-sh	‘look-GER’	qari-di	‘look-3.SG.PAST’

Table 6.2: Examples of *a* vowel reduction.

apet	‘disaster’	apit-i	‘disaster-3.SG.POS’
méwe	‘fruit’	méwi-si	‘fruit-3.SG.POS’
sözle-sh	‘talk-GER’	sözli-di	‘talk-3.SG.PAST’
küche-sh	‘strive-GER’	küchi-di	‘strive-3.SG.PAST’

Table 6.3: Examples of *e* vowel reduction.

Although vowel reduction is generally described as a word-level phenomenon (i.e., word-medial open syllables raise), but raising may be observed in word-final syllables in certain contexts. For example, in utterances like *Adil Hesenge berdi* ‘Adil gave it to Hesen’, the dative suffix ‘-ge’ may raise, producing *Adil Hensengi berdi*.⁴ Our data suggests that the domain for raising is the AP: that is, vowel raising may occur at prosodic word boundaries, but not at AP boundaries. Thus vowel raising may be used as a diagnostic tool for prosodic grouping.

In Fig. 6.7, the final vowel in *nérwa* raises, suggesting that the right edge of this word is not aligned with the boundary of an AP or higher constituent. This is particularly noticeable when this vowel is compared to the [a] in *ram*, which cannot raise because it is in a closed syllable. In this token, the average F1 value for the vowel in *ram* is 654 Hz, while the average F1 for the final vowel in *nérwa* is 338 Hz.

A second diagnostic involves vowel hiatus resolution, which requires resolution across prosodic words within an AP, but does not occur across APs. This is clearly demonstrated by the juncture between *alma* “apple” and *ewetti* “sent” in Fig. 6.8, where the final vowel of *alma* is entirely deleted. Note that this vowel deletion significantly limits the temporal window in which the final H tone of the prosodic word containing *alma* may be realized.

We can thus diagnose the difference between prosodic word boundaries and AP boundaries by looking at not only the height of the H peak at the end of the phrase, but also by whether vowel raising or vowel hiatus resolution takes place between two words.

⁴Note that word-medial raising is represented orthographically, while word-final raising is not.

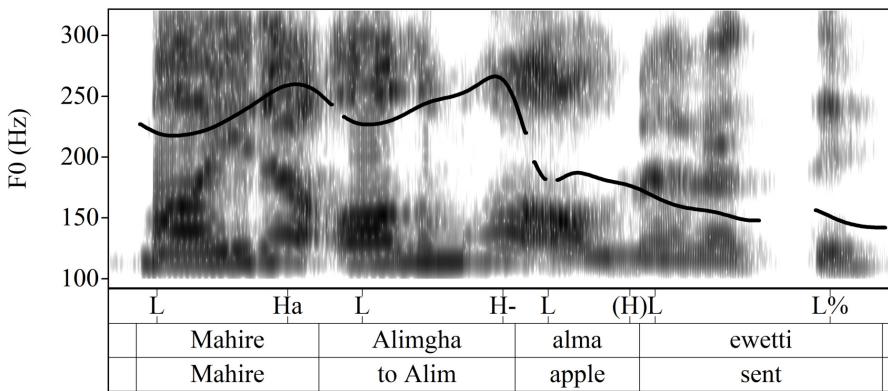


Figure 6.8: Pitch track, spectrogram, and annotation for the sentence *Mahire Alimgha alma ewetti* “Mahire sent Alim an apple.” Note that deletion of the final vowel in *alma* ‘apple’ also results in a severely reduced H tone at the right edge of the corresponding prosodic word.

6.4.4 The accentual phrase and the intermediate phrase

The previous section introduced the AP and compared its properties with those of the next lowest prosodic constituent, the prosodic word. This section will introduce the next highest level on the prosodic tree shown in Fig. 6.6, the *intermediate phrase* (ip) and compare its properties to those of the AP.

Like the AP, the ip has a high tone associated with its right edge, which we write as H-. The most salient differences between this boundary tone and the AP boundary tone Ha is that H- tones are generally higher pitched, and they break declination. Note that in Fig. 6.8 (and in most of the figures shown so far), there is a linear decline in maximum F0 across the utterance, with Ha tones produced later in the utterance displaying lower F0 than those produced earlier. In general, H and Ha tones will show steady declination across an utterance. H- tones behave differently. These do not cause a proper pitch reset (see, e.g., Fig. 6.8, where the overall linear declination trend continues despite the H-), but they do cause a temporary break in declination, reaching F0 peaks similar to or greater than those of the first high tone in the utterance.

Fig. 6.9 shows the affirmative answer to the question “Did Ziya slowly give Alim a wild apple in the bazaar?” This answer is *given*: that is, the entire utterance has already been introduced into the discourse, which means that the utterance does not contain a focused constituent.⁵ Notice that after the first noun phrase *Ziya*, the F0 peaks on *bazarda* “in the bazaar” and *Alimgha* “to Alim” display descending high boundary tones. This declination is one diagnostic we use for

⁵ A given constituent is one that has been mentioned previously in the discourse. To give a clear example from English, consider the following discourse (from Wagner, 2012):

Q: Smith walked into a store. What happened next?

A: A detective arrested Smith

In the second sentence, the discourse referent “Smith” has already been introduced into the discourse. “Smith” is thus considered *given* in the second sentence.

accentual phrases. There is a break in declination following *Alimgha*, where the peak F0 on the adverb *asta* “slowly” is considerably higher than the peak on *Alimgha*. Following *asta* “slowly”, there is declination across the next two words, *yawa* “wild” and *alma* “apple”, which leads into the end of the utterance, which is marked with an IP-final L% tone. Because this sentence lacks a focused element, it illustrates the natural tendency for a pitch reset to break up a long utterance into phrases.

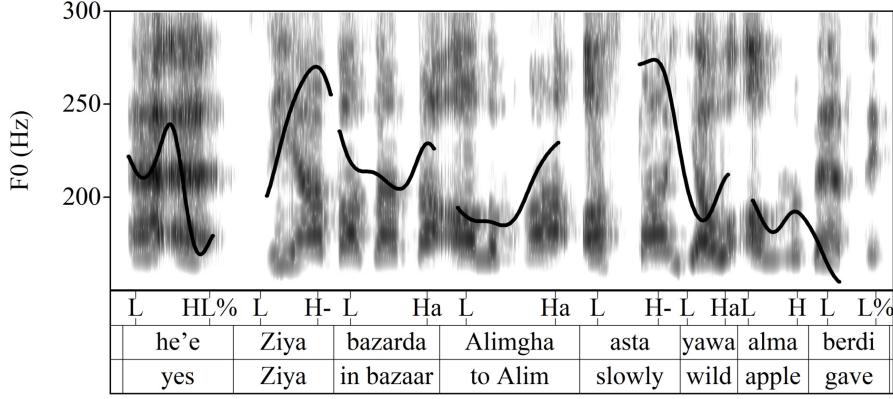


Figure 6.9: Pitch track, spectrogram, and annotation for the sentence *He'e, Ziya bazaarda Alimgha asta yawa alma berdi* “Yes, Ziya slowly gave Alim a wild apple at the bazaar.”

It is generally the case that sentences contain at least two ips. In many simple sentences, the subject forms one ip, while the verb phrase makes up the other. However, it is not always the case that ips correspond to syntactic constituency, as longer utterances often involve divisions where ips do not correspond to a syntactic constituent, as shown in Figs. 6.10 and 6.11. In the former, the subject *momay* “grandmother” is followed by the temporal adverb *bügün* “today”, which bears a higher F0 peak, marking the ip boundary. Based on standard syntactic assumptions, temporal adverbs and the subjects do not form a unique constituent independent of the rest of the utterance.

Similarly, in Fig. 6.11, the initial ip is composed of three APs, the temporal adverb *bügün* “today”, the subject *Meryem*, and the indirect object *Alimgha* “to Alim”, which breaks declination, one of the hallmarks of an ip.

In longer neutral contexts, there seems to be considerable optionality as it relates to prosodic phrasing and in our investigation, we have yet to find an ip that consists of more than three accentual phrases. In non-neutral contexts, the options for prosodic phrasing are more rigid, and are largely determined by information structure, which we address in Section 6.5.1.

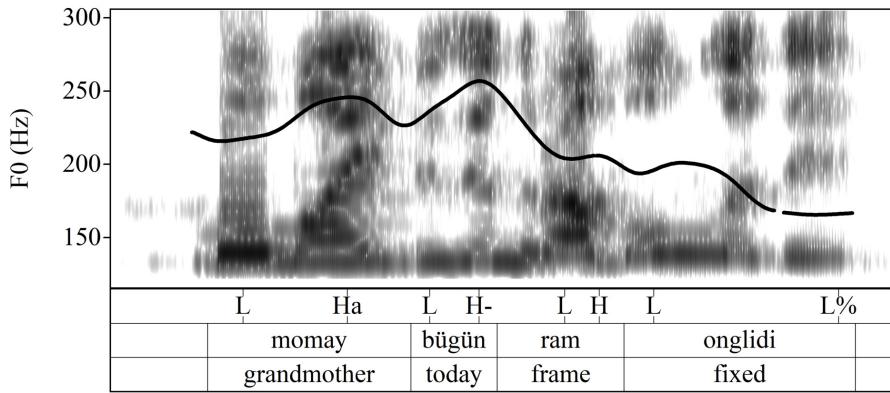


Figure 6.10: Pitch track, spectrogram, and annotation for the sentence *Momay bugün ram onglidi* ‘The grandmother fixed a frame today.’

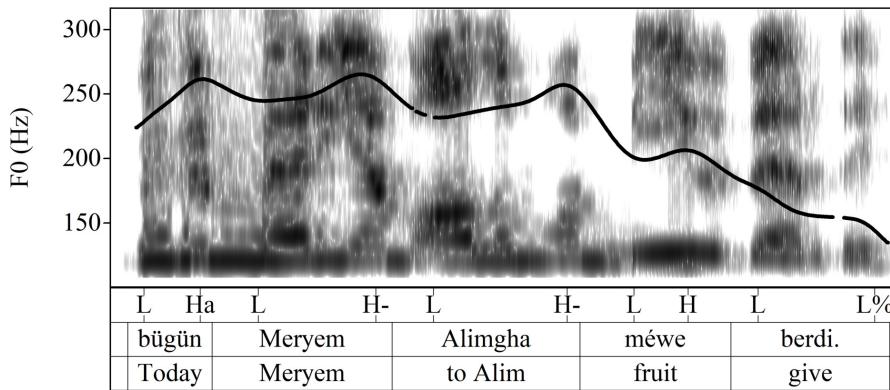


Figure 6.11: Pitch track, spectrogram, and annotation for the sentence *Bügün Meryem Alimgha méwe berdi* “Meryem gave Alim fruit today.”

6.4.5 Disambiguating APs and ips utterance-initially

As indicated in the annotation for Fig. 6.7, it can be difficult to determine whether an initial subject constituent should be labeled as an AP or an ip (and hence its boundary tone labeled as Ha or H-). This is because the most salient difference between these two boundary tones is the triggering of pitch reset by an H- tone. In neutral contexts, the constituent forming the subject generally bears the highest F0 peak, which would be predicted regardless of whether its right edge is marked with an H- or an Ha tone. A clear example of this issue is presented in Fig. 6.12, for which the right edge of the subject, *Adil*, is both the first constituent and bears the highest F0 peak in the utterance.

Despite this ambiguity, we suspect that these initial subject constituents are best treated as ips rather than APs. The intuition behind this is based on a tendency for native speakers to perceive the juncture between subject and object in such sentences as greater than the juncture between object and verb.

To provide empirical support for this intuition, we carried out a small phonetic study comparing simple sentences like *Qurban mangidu* “*Qurban* will walk” (Fig. 6.13) against more complex sentences like *Qurban bilen Ziba mangidu* “*Ziba* will walk with *Qurban*” (Fig. 6.14) or *Qurban bilidighan Ziba ketti* “*Ziba*, who knows *Qurban*, arrived”. Although the complex sentences begin with the same word as the corresponding simple sentence, their syntactic structure differs: in the first example given above, for example, *Qurban* is the complement of the postposition *bilen*.⁶

We predicted that the constituency of these noun phrases should differ between these simple and complex sentences: specifically, that the subject of the simple sentences should form an H-, in line with speaker intuitions about juncture size, while the initial noun phrase in the complex sentences should form an AP. We predicted that these differences in constituency would be reflected in the pitch and duration of the final syllables of the target noun phrases.

We asked six native Uyghur speakers to read a set of 18 sentences with the following three question and response structures:

- **Simple:**

Néme boldi? ‘What happened?’
NAME VERB ‘NAME VERB’

- **Complex comitative:**

Néme boldi? ‘What happened?’
NAME₁ bilen NAME₂ VERB ‘NAME₂ VERB with NAME₁’

- **Complex relative clause:**

Néme boldi? ‘What happened?’
NAME₁ bilidighan NAME₂ ketti ‘NAME₂, who knows NAME₁, arrived.’

The values for NAME were *Ziba* or *Qurban*⁷ and the values for VERB were *yürudu* “will go”, *mangidu* “will walk”, *bardi* “went”, or *yighlaptu* “wept.” The tense changes across verbs were made to prevent speakers from getting bored and falling into list intonation.

⁶Note that the pitch peak on the end of *Qurban* in Fig. 6.14 suggests that *bilen* is indeed a postposition rather than a suffix, like *-gha* or *-ni* (cf. Figs. 6.8 and 6.12).

⁷Due to researcher error, *Adil* was erroneously used instead of *Qurban* in the complex relative clause construction by two speakers.

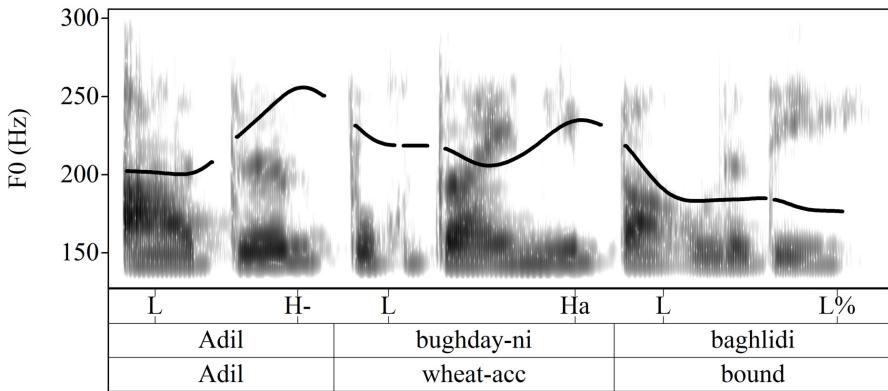


Figure 6.12: Pitch track, spectrogram, and annotation for the sentence *Adil bughdayni baghlidi* “Adil bound the wheat.”

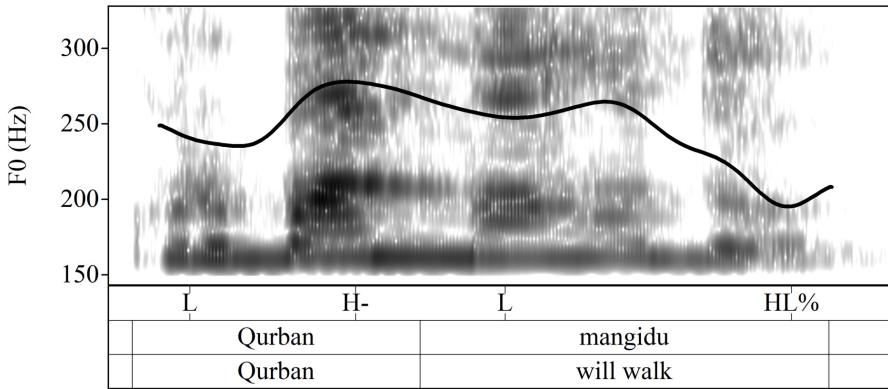


Figure 6.13: Pitch track, spectrogram, and annotation for the sentence *Qurban mangidu* “Qurban will walk.”

The final syllables of the initial word in each of these utterances were segmented using Praat. We extracted duration, mean F0, max F0, and mean intensity for each syllable. As in the stress study presented in Section 6.3, we fit linear mixed effects models for each of these measurements. The independent variable was subjecthood: whether the name served as a subject (simple sentences) or as a non-subject (complex sentences). The duration model also included an independent variable for whether the syllable had a coda (as in *Qurban*) or not (as in *Ziba*). We included random intercepts for each participant and random slopes for each participant over subjecthood.

The duration model showed a significant effect of subjecthood on duration: the final syllable of subjects tends to be longer than that of non-subjects in the same position (Fig. 6.15; $\beta = 0.013, t = 2.461, p < 0.05$). Unsurprisingly, there was also a significant effect of coda presence

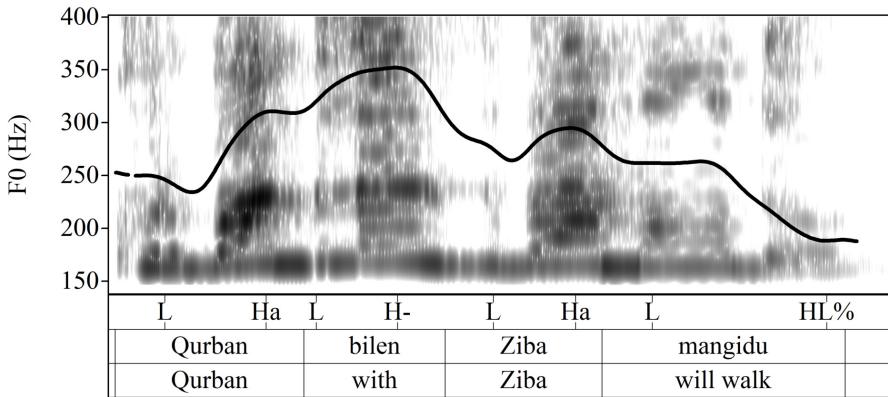


Figure 6.14: Pitch track, spectrogram, and annotation for the sentence *Qurban bilen Ziba mangidu* “Ziba will walk with Qurban.”

on duration ($\beta = 0.037, t = 7.306, p < 0.001$). None of the models of F0 or intensity showed significant differences based on subjecthood.

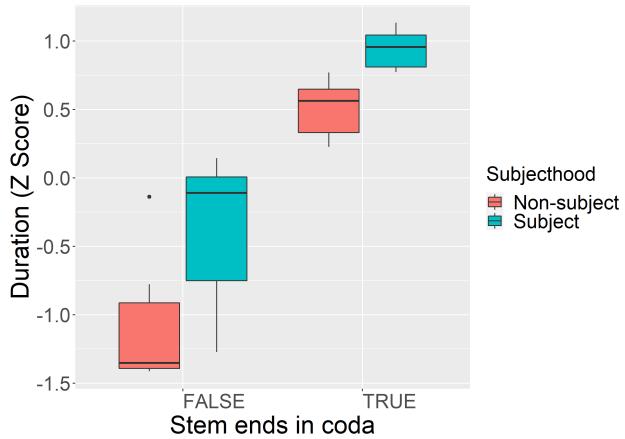


Figure 6.15: Comparison of duration of final syllables between subjects and non-subjects in word-initial position. Durations are normalized to Z scores to facilitate comparison across participants.

We interpret these differences in duration of final syllables between word-initial subjects and word-initial non-subjects to reflect a difference in their constituency. Research has shown that larger prosodic boundaries are associated with a longer duration on the last syllable of the prosodic unit (e.g., Klatt, 1975; Lehiste, Olive, & Streeter, 1976; Wightman, Shattuck-Hufnagel, Ostendorf, & Price, 1992), and the systematic difference in length between the two positions here supports different boundary strengths. Though this study does not provide a method to resolve the ambiguity between APs and ips in utterance-initial subjects on a case-by-case basis, it suggests they should generally be treated as ips.

6.4.6 The intermediate phrase and the intonational phrase

Like ips, *intonational phrases* (IP) are marked by a boundary tone on their final syllable. This syllable is also substantially lengthened and generally followed by a pause. The final word of an IP, usually the final verb of an utterance, bears one of four boundary tones: L%, H%, HL%, or LH%. Note that these are the only cases of a complex tones in our model.

Generally speaking, L% and HL% mark the end of declarative utterances while H% and LH% mark the end of polar and wh-questions, emphatic sentences, and serve as an indicator that the speaker plans to continue speaking on the same topic. All figures presented up to this point include examples of declarative L% or HL% tones. An example of an H% tone is provided in Fig. 6.16, which represents a simple yes/no question. In this case, the H% target is hosted by the final question particle *mu*. The pragmatic distinction between the simple and corresponding complex tones is not clear: we return to this issue in the discussion.

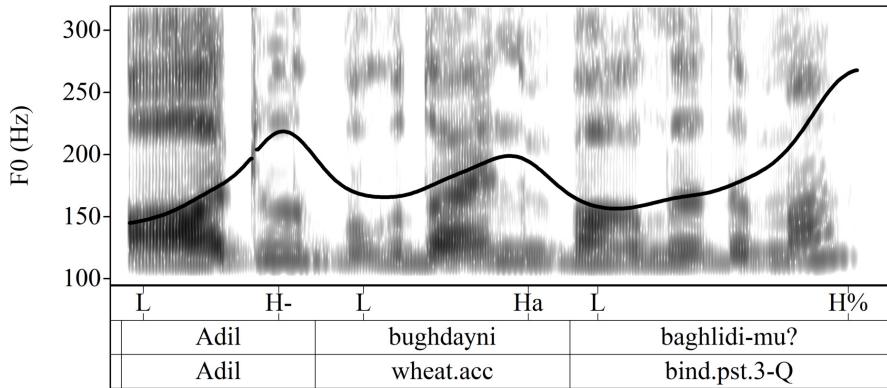


Figure 6.16: Pitch track, spectrogram, and annotation for the sentence *Adil bughdayni baghlidi-mu?* “Did Adil bind the wheat?”

We will postpone further discussion and examples of IP-final tones to the next section, where we will discuss the intonational properties of various utterance types.

6.5 The intonational properties of various utterance types

The previous section provided an overview of the core aspects of our intonational model. In this section, we will present the characteristic intonational patterns of a range of different utterance types.

6.5.1 Focus and questions

Because Uyghur lacks head-marking tones, prominent elements must be indicated prosodically using boundary tones. The most common prosodic strategy to indicate prominence in Uyghur is

to align the focused constituent with the right edge of an ip, or, equivalently, produce an H- tone on the focused constituent's right edge. Recall the simple sentence *Adil bughdayni baghlidi* “Adil bound the wheat” in Fig. 6.12, which represents prototypical Uyghur prosody: the subject *Adil* displayed the highest F0 peak in the utterance, followed by a slightly lower F0 peak on the object, *bughdayni* “wheat”, which bears the hallmarks of an Ha tone, followed by the sentence-final declarative L%. We can compare this with the same sentence but with the additive/focus clitic =mu “even/also” added to the object, as in Fig. 6.17. Here the focus on the object is indicated by its promotion to the final element of an ip constituent. The boundary tone occurring at its right edge displays the declination break and high F0 peak that are characteristic of H- tones.

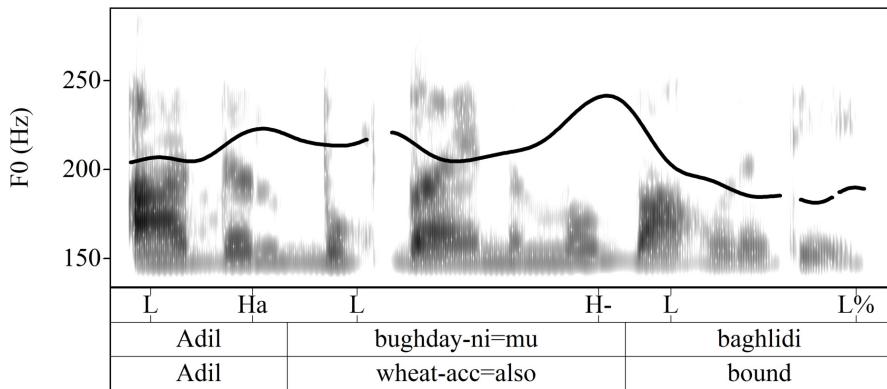


Figure 6.17: Pitch track, spectrogram, and annotation for the sentence *Adil bughdaynimu baghlidi* “Adil even bound the wheat.”

The prosody of focus when there is no focus particle is almost identical. For instance, wh-expressions generally receive the same kind of intonational prominence as other types of focused NPs, as demonstrated in Fig. 6.18.

Like the polar question shown in Fig. 6.16, the IP in Fig. 6.18 also ends in an H% tone. In addition to this final tone, the wh-expression *kimge* “to whom” is focused and is thus located at the right edge of an ip, bearing an H- tone. The preceding material, the subject *Amine* and the temporal adverbial *diishenbe* “Monday” form an ip of their own.

A response to this question, shown in Fig. 6.19 has almost the same intonational properties. The sole difference is the final L% tone, which indicates a declarative sentence. *Meryemge* “to Meryem” is the only new information introduced in the answer, and thus receives focus by occurring at the end of an ip. The material preceding *Meryemge* forms an independent ip, as was the case for the question word in Fig. 6.18.

Corrective focus gets the same prosodic treatment as answers to questions. Consider the adverb *asta* “slowly” in Fig. 6.20. The context is one in which the interlocutor suggests that the event was carried out quickly. The speaker accepts the entire sentence except for the manner adverbial. As a result, the speaker offers a correction to the appropriate manner adverbial, replacing “quickly” with “slowly”, and making “slowly” the most prominent word in the sentence. This prominence is indicated by the fact that the final syllable bears an H- tone, which clearly breaks declination

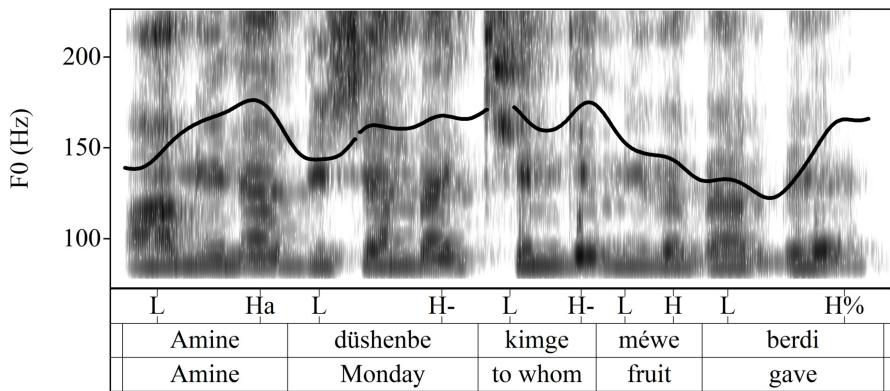


Figure 6.18: Pitch track, spectrogram, and annotation for the sentence *Amine düshenbe kimge méwe berdi*? “Who did Amine give fruit to on Monday?”

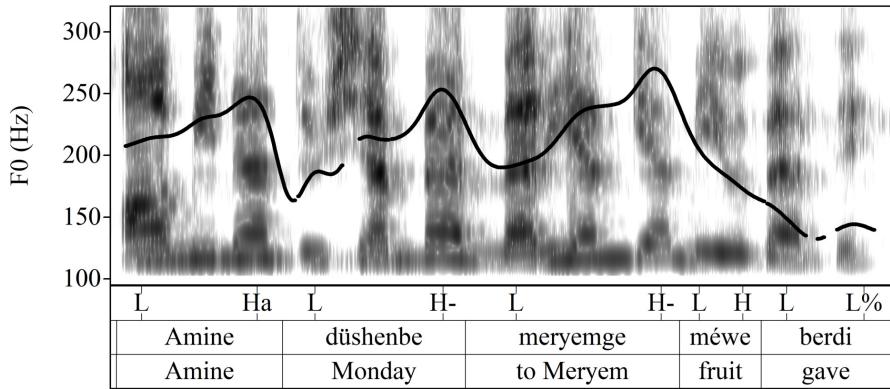


Figure 6.19: Pitch track, spectrogram, and annotation for the sentence *Amine düshenbe Meryemge méwe berdi* “Amine gave fruit to Meryem on Monday.”

when compared to the preceding peak on *Alimgha* “to Alim”. In neutral contexts, adverbs generally constitute prosodic words or APs,

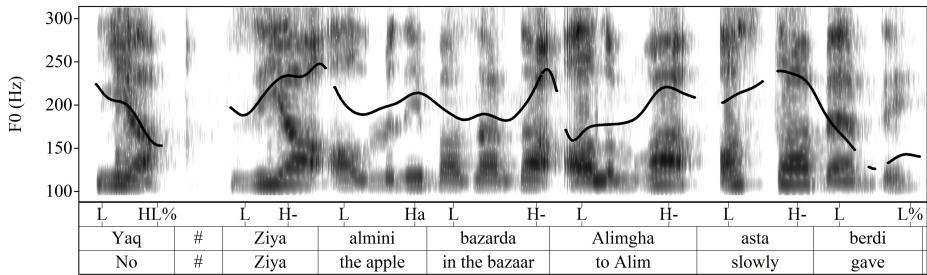


Figure 6.20: Pitch track, spectrogram, and annotation for the sentence *Yaq, Ziya almini bazarda Alimgha asta berdi* “No, Ziya slowly gave the apple to Alim at the bazaar.”

6.5.2 Discourse and turn taking

In addition to marking questions, IP-final H% and LH% can also be used to mark a continuation rise, indicating that the speaker intends to continue speaking. For instance, the utterance represented by Fig. 6.21 is a response to the question “Where is the apple?”. The precise context is one in which there is an apple in the common ground of the two interlocutors, but the questioner cannot find it. The first sentence *Alma yoq* “The apple is gone” bears a H% tone on its right edge, which indicates that the speaker has not yet finished speaking. The second part of the utterance *Almini Ziya yédi* bears normal declarative intonation. Because the apple is the topic in the second sentence, it occurs sentence-initially and gets the H- that is generally associated in the initial constituent in discourse neutral contexts.

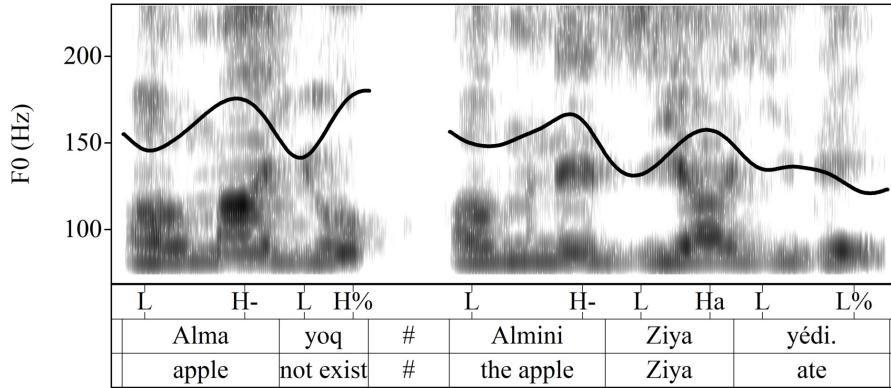


Figure 6.21: Pitch track, spectrogram, and annotation for the sentence *Alma yoq. Almini Ziya yédi*. “The apple is gone. Ziya ate it.”

This use of H% or LH% is also very common when one gives orders or instructions; in particular, when the order or instructions consist of more than a single sentence or at least of more than a single clause. This is exemplified by the naturalistic utterance in Fig. 6.22.

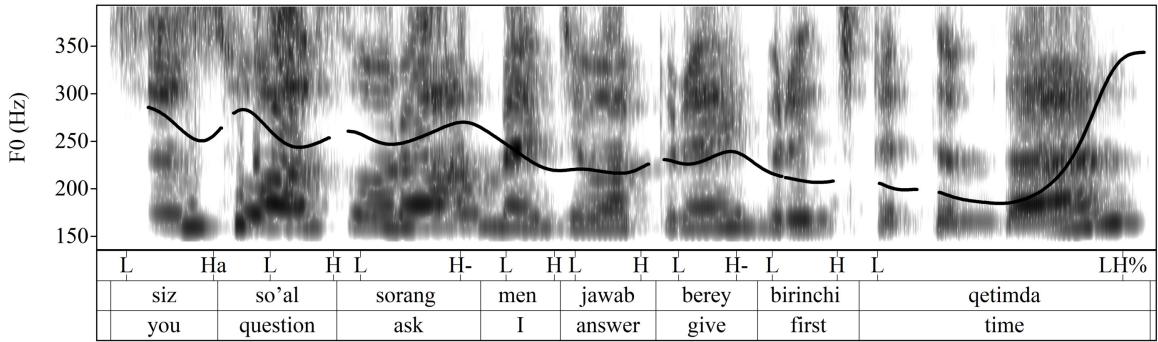


Figure 6.22: Pitch track, spectrogram, and annotation for the sentence *Siz so'al sorang, men jawab béréy, birinchi qéttimda* “You ask a question, I'll give an answer, for the first time.”

In this particular case, the speaker is giving instructions to her interlocutor about how they will read our stimuli. She indicates that the first time, the interlocutor will be the one reading the questions and she will answer. The H% boundary tone at the end of this utterance indicates that she has not finished providing instructions, and that her interlocutor should not interrupt. She goes on to say that the second time, the questioner and responder roles will be reversed (this portion of the utterance is not shown).

In addition to instructional contexts, use of H% in non-interrogative contexts is extremely common in narratives, where the speaker plans to speak uninterrupted until they are finished. Consider Fig. 6.23, which shows the final two sentences of a response to an interview question about the speaker's experiences as a teacher.

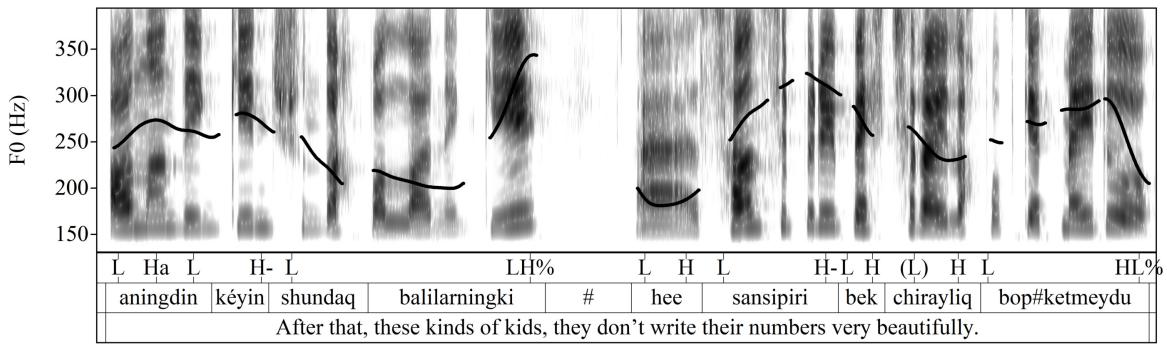


Figure 6.23: Pitch track, spectrogram, and annotation for the sentence *Aningdin kékyn shundaq balilarningki, hee, sansipiri bek chirayliq bop ketmeydu* “After that, these kinds of kids, well, they don't write their numbers very beautifully.”

The first sentence is the penultimate utterance in the chain, and ends with an LH% tone, indicating that the speaker is not yet relinquishing her turn. The final sentence bears the declarative HL%,

indicating that she has finished speaking.

It is not always the case that H% boundaries associate with the right edges of complete sentences. For instance, the converbial, clause-chaining suffix *-ip* also often carries an H% tone in conjunction with a substantial pause. This is exemplified by Fig. 6.24, which consists of a sequence of three clauses in a single utterance, the first two inflected by *-ip* and the last one fully finite. The first two clauses bear sharp, rising H% tones, while the final clause does not.

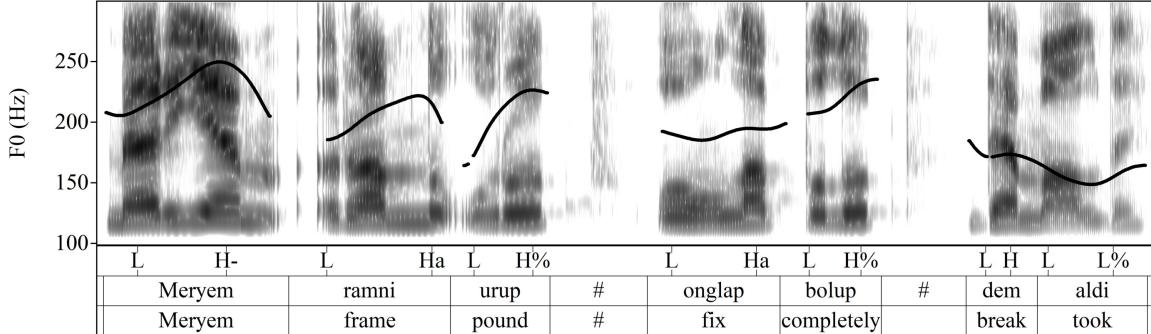


Figure 6.24: Pitch track, spectrogram, and annotation for the sentence *Meryem ramni urup, onglap bolup, dem aldi* “Meryem pounded the frame, fixed it completely, and then took a break.”

6.6 Discussion

The model proposed in this chapter is intended as a significant, but still somewhat preliminary, step toward a comprehensive model of Uyghur intonational phonology. Moving forward, there are various ways in which the model can be expanded upon and improved. Investigating more naturalistic speech and varying the types of discourse is one important step in this process. The model as presented here offers a framework for investigating such complex data.

Furthermore, there are many constructions that bear interesting intonational properties and require independent investigations. There seem to be differences that vary with tense/evidentiality, agreement, clitics vs. suffixes, clausal embedding, and different types of mood marking. Even in the context of focus, our model shows that focused constituents form independent ips, but more research is necessary to determine the properties of non-focused constituents in constructions that contain focus (see for example, the deaccenting that appears to take place following the focused word in Figs. 6.1 and 6.2).

Despite our model having been created and validated based on a limited set of constructions elicited specifically for this purpose, however, we are confident that it is flexible enough as it currently stands to robustly describe the majority of intonational patterns encountered in Uyghur.

Take for instance the long naturalistic utterance presented in Fig. 6.25, which is from a broadcast on Radio Free Asia (Hoshur, 2010) and contains constructions that we have not discussed above. The first IP is compatible with the model as presented. The subject *u* “s/he” bears an H- on its right edge, which shows declination into the right edge of the following NP *so’al-jawab* “Q&A”,

which is marked with an Ha tone. This clause is not the end of the broadcaster's speech, which results in it bearing an H% that indicates he intends to continue speaking. The second clause contains a nominalized complement clause embedded under the verb "explain", which also contains an embedded question *nême üchün* "why". Despite the fact that we did not explicitly discuss embedded clauses, there are no tonal patterns present that are incompatible with those discussed in this chapter (though the reason for the elision of the final two L tones is not clear).

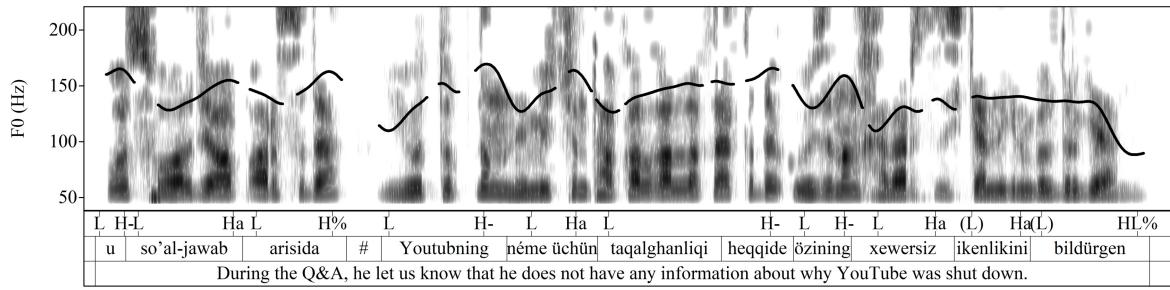


Figure 6.25: Pitch track, spectrogram, and annotation for a complex, naturalistic utterance from Radio Free Asia.

One area of the model that requires additional investigation is the difference between the simple (L% and H%) and complex (HL% and LH%) IP boundary tones. The complex tones are generally used in similar circumstances to their simple counterparts: that is L% and HL% mark declarative utterances while H% and LH% mark interrogatives and continuation rises. It is not clear from our elicitation whether there are significant semantic or pragmatic distinctions between these pairs: our speakers generally treat both as acceptable in the same contexts. Subsequent production and perception studies will be valuable for understanding the distribution and function of these IP-final tones.

We have also observed, especially among Kazakh speakers, a complex HL% contour at the ends of certain questions, as shown in Fig. 6.26. The general pattern found for questions involves either a gradual (H%) or sharp (LH%) a sharp rise towards the final syllable of the utterances in questions. It is unclear what function the HL% contour is serving in this case.

Finally, although we suggest that prosodic words generally consist of a single stem and its affixes, we have come across cases where multi-word expressions appear to form a single prosodic word: for example, in Fig. 6.25, the expression *taqalghanliqi heqqide* "about (its) having been shut down" appears to phrase a single prosodic word. In addition to post-positional phrases like this one, we have also observed this patterning in relative clause constructions, where the head of the relative clause will often phrase as a single prosodic word with the main verb of the relative clause. We leave this, and other topics, as areas for future research.

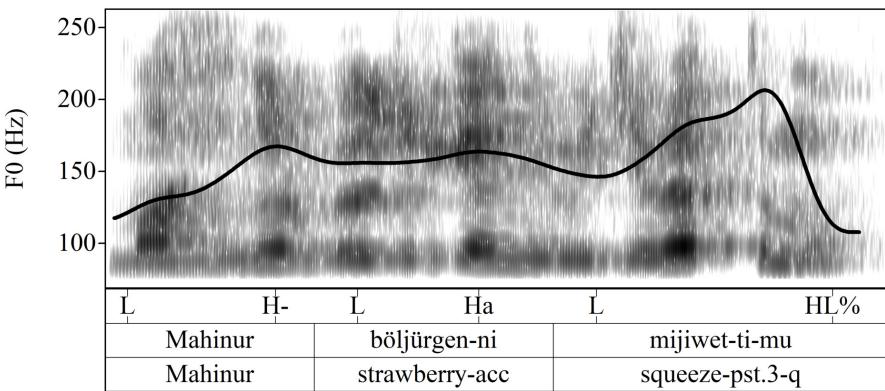


Figure 6.26: Pitch track, spectrogram, and annotation for a polar question ending in an HL% contour.

6.7 Conclusions

In this chapter, we have presented a preliminary model of the intonational phonology of Uyghur. We have shown that duration is the most reliable indicator of prominence in Uyghur words, in line with (Yakup, 2013). Furthermore, pitch does not associate with the prominent syllable, i.e. the head, but rather with the edges of phrases. As such, Uyghur appears to be a language with prosodic heads that are ignored by the intonational system. We then motivated the following prosodic constituents in Uyghur: the prosodic word, accentual phrase, intermediate phrase, and intonational phrase. Prosodic words begin with a low tone, which results in all higher level prosodic constituents beginning with a low tone. We then showed that the right edges of prosodic words and APs are marked with a high tone, but show successive declination until an IP boundary is met. Right edges of IPs are similarly marked with a higher tone relative to APs and break declination when following an AP. Finally, IPs are the largest prosodic constituent in our model, which indicate higher level discourse functions, such as sentence type (e.g. declarative vs. interrogative) or other discourse information.

We are optimistic that this model will help facilitate further study of Uyghur intonation, particularly as it interfaces with syntax. We have already used a version of this model to explore how intonation can be used to diagnose direct quotation vs. indexical shifted readings of embedded clauses (Major & Mayer, 2019), and we hope that it will be useful for other researchers studying Uyghur prosody and phonology, as well as for theories of intonational typology.

6.8 Acknowledgements and notes

We would like to thank our consultants for sharing their language and culture with us: Ziba Ablet, Mustafa Aksu, Akbar Akmat, Gülnar Eziz, Abduquyym Mamat, Memet Semet, Mahire Yakup, and the students of School 153 in Almaty, Kazakhstan. Without their generosity and time, none of this would be possible. We would also like to thank Sun-Ah Jun, the attendees of Speech Prosody

9 and the Intonational Phonology of Typologically Rare or Understudied Languages, and the attendees of the UCLA Phonetics Seminar for their invaluable feedback. The authors of this paper are listed in alphabetical order.

References

- Barker, C. (1989). Extrametricality, the cycle, and Turkish word stress. *Phonology at Santa Cruz*, 1, 1-34.
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1–48. doi: 10.18637/jss.v067.i01
- Beckman, M., & Pierrehumbert, J. (1986). Intonational structure in Japanese and English. *Phonology Yearbook*, 3, 255-309.
- Boersma, P., & Weenink, D. (n.d.). *Praat: doing phonetics by computer [Computer Program]*. Version 6.0.37. <http://www.praat.org/>. (Accessed: 2018-02-17)
- Dobrovolsky, M. (1999). The phonetics of Chuvash stress: Implications for phonology. In *Proceedings of the 14th International Congress of Phonetic Sciences*, 1, 539-542.
- Engesæth, T., Yakup, M., & Dwyer, A. (2009/2010). *Teklimakandin salam: hazirqi zaman Uyghur tili qollanmisi / greetings from the Teklimakan: a handbook of modern Uyghur*. Lawrence: University of Kansas Scholarworks.
- Günes, G. (2015). *Deriving prosodic structures*. LOT, Netherlands Graduate School.
- Hahn, R. F. (1991a). Diachronic aspects of regular disharmony in modern Uyghur. In W. Boltz & M. Shapiro (Eds.), *Studies in the Historical Phonology of Asian Languages*. John Benjamins.
- Hahn, R. F. (1991b). *Spoken Uyghur*. Seattle, WA: University of Washington Press.
- Hoshur, S. (2010). Google plans to leave china. *Radio Free Asia*.
- Inkelas, S. (1999). Exceptional stress-attracting suffixes in Turkish: representations versus the grammar. *The prosody-morphology interface*, 134, 187.
- Inkelas, S., & Orgun, C. O. (1998). Level (non) ordering in recursive morphology: evidence from Turkish. In P. Farrell (Ed.), *Morphology and its relation to phonology and syntax* (p. 360-392).
- Ipek, C. (2015). *The phonology and phonetics of Turkish intonation* (Unpublished doctoral dissertation). University of Southern California.
- Ipek, C., & Jun, S.-A. (2013). Towards a model of intonational phonology of Turkish: Neutral intonation. In *Proceedings of meetings on acoustics ica2013* (Vol. 19, p. 060230).
- Jun, S.-A. (2005). Prosodic typology. In S.-A. Jun (Ed.), *Prosodic typology: The phonology of intonation and phrasing* (p. 430-458). Oxford: Oxford University Press.
- Kabak, B., & Vogel, I. (2001). The phonological word and stress assignment in Turkish. *Phonology*, 18(3), 315–360.
- Kaisse, E. (1985). Some theoretical consequences of stress rules in Turkish. *Proceedings of the 21st meeting of the Chicago Linguistic Society*, 1, 199-209.
- Kamali, B. (2011). *Topics at the PF interface of Turkish* (Unpublished doctoral dissertation). Harvard University.
- Kan, S. (2009). *Prosodic domains and the syntax-prosody mapping in Turkish* (Unpublished master's thesis). Boğaziçi University.

- Kisseberth, C. W., & Abasheikh, M. I. (2011). Chimwiini phonological phrasing revisited. *Lingua*, 121, 1987–2013.
- Klatt, D. H. (1975). Vowel lengthening is syntactically determined in a connected discourse. *Journal of Phonetics*, 3(3), 129–140.
- Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, 82(13), 1–26. doi: 10.18637/jss.v082.i13
- Ladd, D. R. (1996/2008). *Intonational phonology* (1st and 2nd ed.). Cambridge: Cambridge University Press.
- Lees, R. (1961). The phonology of Modern Standard Turkish. In *Uralic and altaic series* (Vol. 6). Indiana University Publications.
- Lehiste, I., Olive, J. P., & Streeter, L. A. (1976). Role of duration in disambiguating syntactically ambiguous sentences. *The Journal of the Acoustical Society of America*, 60(5), 1199–1202.
- Levi, S. V. (2005). Acoustic correlates of lexical accent in Turkish. *Journal of the International Phonetic Association*, 35(1), 73–97.
- Lindström, E., & Remijsen, B. (2005). Aspects of the prosody of Kuot, a language where intonation ignores stress. *Linguistics*, 43(4), 839–870.
- Major, T., & Mayer, C. (2018). Towards a phonological model of Uyghur intonation. In K. Klessa, J. Bachan, A. Wagner, M. Karpiński, & D. Śledziński (Eds.), *Proceedings of the 9th speech prosody international conference*.
- Major, T., & Mayer, C. (2019). What indexical shift sounds like: Uyghur intonation and interpreting speech reports. In M. Baird & J. Pesetsky (Eds.), *NELS 49: Proceedings of the forty-ninth annual meeting of the north east linguistic society* (Vol. 2, pp. 255–264).
- Mayer, C. (2020). *Gradient opacity in Uyghur backness harmony*.
- Menges, K. (1995). *The Turkic languages and peoples* (2nd ed.). Harassowitz Verlag.
- Özçelik, Ö. (2015). Stress or intonational prominence? Word accent in Kazakh and Uyghur. *Turkic Languages*, 19, 163–192.
- Pierrehumbert, J. (1980). *The phonology and phonetics of English intonation* (Unpublished doctoral dissertation). MIT.
- R Core Team. (2017). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria. Retrieved from <https://www.R-project.org/>
- Wagner, M. (2012). Focus and givenness: A unified approach. *Contrasts and positions in information structure*, 102–147.
- Wightman, C. W., Shattuck-Hufnagel, S., Ostendorf, M., & Price, P. J. (1992). Segmental durations in the vicinity of prosodic phrase boundaries. *The Journal of the Acoustical Society of America*, 91(3), 1707–1717.
- Yakup, M. (2013). *Acoustic correlates of lexical stress in native speakers of Uyghur and L2 learners* (Unpublished doctoral dissertation). University of Kansas.