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, which is typologically unusual. 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 present 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, which is typologically unusual. 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. All data and recordings used in this chapter can be found at `https://github.com/connormayer/uyghur_into`

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 recently 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) and (Ipek, 2015) 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 prosody

Throughout this paper, we will indicate stressed syllables using capitalization. We use this notation because the acute accent conventionally used to mark stress indicates a phonemic contrast in Uyghur Latin orthography: the character *é* represents the sound /e/, while *e* represents /æ/. Assuming that each word has some stressed syllable, we do not mark stress on monosyllabic words. Stress location is primarily based on Engesæth et al. (2009/2010), which is the only Uyghur pedagogical material we are aware of that explicitly marks stress. For words that are not listed in this resource, we assigned stress according to subjective judgments by one of our consultants.

The status of stress in Uyghur is not well understood. Nadzhip (1971) describes Uyghur stress as being "remarkable for its complexity and instability," (p. 63). He suggests that stress typically falls on the final syllable in the word, but that there are numerous exceptions to this generalization, particularly among loanwords.

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 rightmost 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"). In addition to stress, Hahn (1991b) briefly discusses the use of pitch in Uyghur (p. 29). He claims based on impressionistic data that utterances, including questions, tend to end in low tones, and that focus and question phrases tend to exhibit an expanded pitch range.

Engesæth et al. (2009/2010) agree that Uyghur is 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., *TAPshuruq* 'homework', *muREKkep* 'complicated'), but note exceptions to this (e.g., *turPAN* 'Turpan'). In addition, they suggest that loan words maintain stress patterns from the original language (e.g., *gimNAStika* 'gymnastics' from Russian; *BAla* 'disaster' from Farsi; cf. the native Uyghur word *baLA* 'child'). They also suggest that the primary acoustic correlate of stress is duration, not f0 or intensity.

A series of production and perception experiments by Yakup (2013) supports this claim. These experiments targeted stress minimal pairs or near minimal pairs such as *BAla* 'disaster' and *baLA* '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 sometimes 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 chapter (e.g., intonational prominence on the rightmost edges of words, the presence of idiosyncratic stress), the study has methodological limitations: for example, 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 f0 differences 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]

To our knowledge, the sole description in the AM framework of a language with a stress system that has no intonational marking is Abbas (2021); Abbas and Jun (2021), which describe the intonational phonology of Farasani Arabic. Farasani Arabic is claimed to have lexical stress, but exhibits head-marking intonation only in focus contexts: that is, when a word is focused the pitch peak falls on its stressed syllable, but when it is not focused, the peak always falls on the AP-final syllable.

A number of other languages also appear to have a stress system with no intonational marking,

[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: stress-based analyses are more common in the previous literature; no Uyghur orthography distinguishes between long and short vowels despite being an accurate phonemic representation of the language in most other respects (see Mayer, 2021, Section 2.2); 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.

but these have not been described in the AM framework: 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, 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) presents evidence that stress in Uyghur is reflected only by vowel length, not f0 or intensity. This suggests that although Uyghur can be described as a stress language, stress and intonational tone are independent. In this section, we use acoustic measurements to confirm that vowel duration 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 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.

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 the Xinjiang Uyghur Autonomus Region in the People's Republic of China: three from the greater Urumqi area, which is the capital of Xinjiang located in the northern part of the region, and one from Qashqar, located in southwestern Xinjiang. The other four (2M, 2F) are from the Almaty region in southeastern 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 in Almaty.

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 very 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. Because Uyghur orthography does not mark stress, we disambiguated between stress minimal pairs by inserting Russian (in Kazakhstan) or English (in the United States) translations of the target words when necessary. This design allowed a balanced number of stressed and unstressed vowels in similar contexts and with mostly the same vowel quality. This resulted in 64 vowel tokens per speaker (16 words x 2 vowels per word x 2 sentence contexts). Due to errors in data collection, two speakers from Xinjiang are missing several tokens, for a total of 504 vowel tokens across all speakers.[3]

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 are 80 ms and 118 ms.[4]

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.

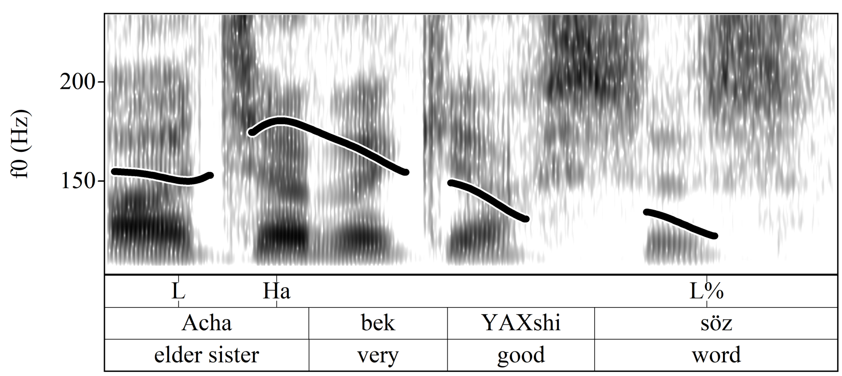

Figure 6.1: Pitch track of the word *Acha* 'elder sister' in sentence-initial position. This sentence means "'Elder sister' is a very good word." The contents of the tone tier are explained in Section 6.4.

[3]One speaker is missing a token of *daDA* in sentence-initial position; a second speaker is missing tokens of *baHA* in sentence-medial position, *DAcha* in sentence-medial position, and *daDA* in sentence-medial position.

[4]Note that the target word in Figs. 6.1 and 6.2 is focused, and the material following the target word is realized as a single phrase with only a final boundary tone, resulting in a fairly linear decline in f0. We return to this phenomenon in Section 6.5.1.

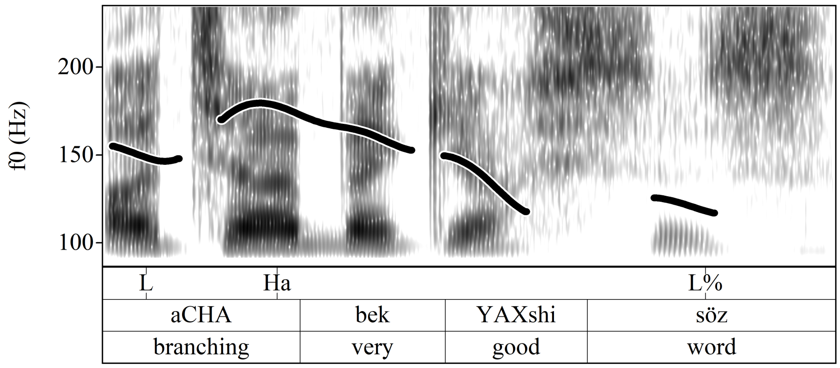

Figure 6.2: Pitch track of the word *aCHA* 'branching' in sentence-initial position. This sentence means "'Branching' is a very good word." The contents of the tone tier are explained in Section 6.4.

6.3.3 Analysis

The two vowels in each target word were segmented using Praat (Boersma & Weenink, 2021), and the mean intensity (in dB), 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 f0 respectively. Our independent variables were stress (stressed or unstressed), position in the word (initial or final syllable), and position of the word in the sentence (initial or medial), as well as an interaction term between position in word and position in sentence. We included random intercepts for word and participants.

6.3.4 Results

The duration model showed a significant main effect of stress ($\beta = 0.016, t = 6.312, p < 0.001$), with stressed vowels being significantly longer; a significant main effect of position in the word ($\beta = 0.013, t = 3.674, p < 0.001$), with vowels in the final syllable of words being significantly longer; and a significant main effect of position in sentence ($\beta = -0.008, t = -2.071, p < 0.05$), with vowels in sentence-medial words being significantly shorter. Duration values are plotted in Fig. 6.3.

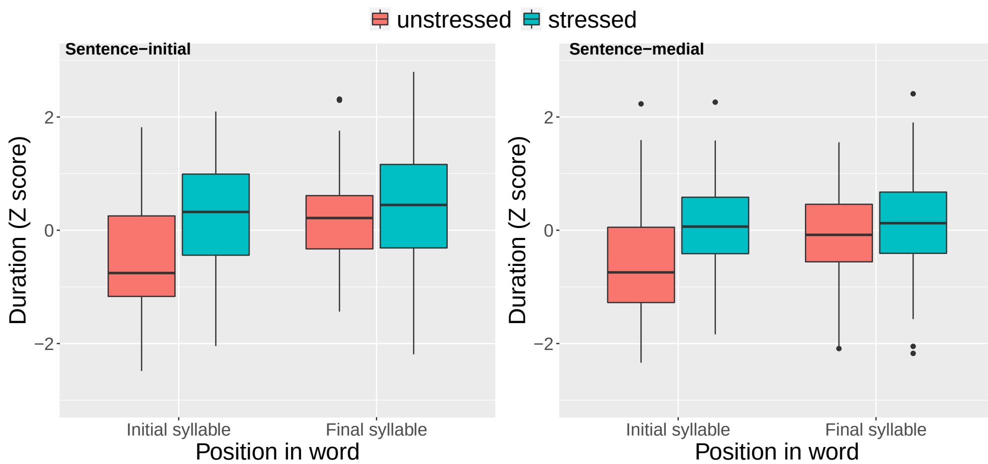

Figure 6.3: Vowel duration broken down by position of word in sentence, position of vowel in word, and whether or not the vowel is stressed. Values are reported in Z scores rather than seconds to facilitate comparison between speakers.

The intensity model showed significant main effects of position in sentence ($\beta = -1.083, t = -3.762, p < 0.001$), with words in sentence-medial position having significantly lower intensity, as well as a significant effect of position in word, with vowels in word-final position having higher intensity ($\beta = 1.079, t = 3.764, p < 0.001$). There was no significant effect of stress on intensity. Intensity values are plotted in Fig. 6.4.

The f0 model showed a significant main effect of position in the word ($\beta = 42.574, t = 18.267, p < 0.001$), with word-final syllables having a significantly higher f0, and a significant main effect of position of the word in the sentence ($\beta = -8.479, t = -3.622, p < 0.001$), with vowels in sentence-medial words having a lower f0, 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 f0 ($\beta = -17.247, t = -5.213, p < 0.001$). We interpret this interaction as capturing the pattern of exponential decay that has been observed for f0 declination in other languages (e.g. Liberman & Pierrehumbert, 1984; Prieto, Shih, & Nibert, 1996). There were no significant effects of stress on f0.

These results show that stress location is a significant predictor of duration, but not f0. f0,

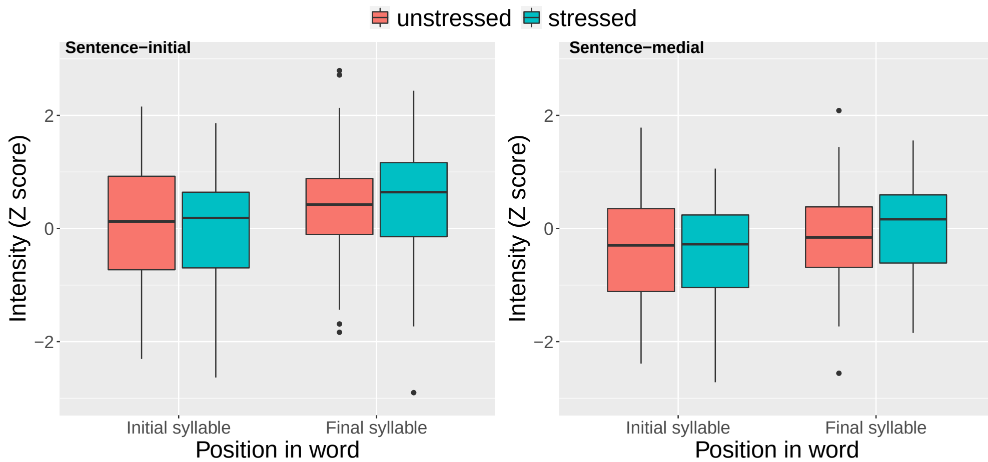

Figure 6.4: Mean vowel intensity broken down by position of word in sentence, position of vowel in word, and whether or not the vowel is stressed. Values are reported in Z scores rather than dB to facilitate comparison between speakers.

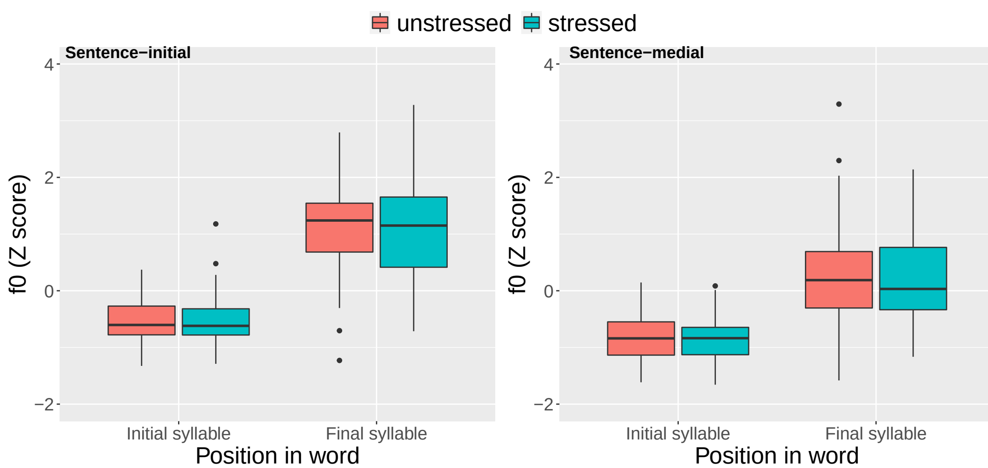

Figure 6.5: Mean vowel f0 broken down by position of word in sentence, position of vowel in word, and whether or not the vowel is stressed. Values are reported in Z scores rather than Hz to facilitate comparison between speakers.

rather, is predicted by the position in the word (word-final syllable $>$ word-initial syllable), reflecting the edge-marking function of pitch, and the position of the word in the utterance, reflecting declination. These results support treating Uyghur as a stress language with only edge-marking intonation. In the remainder of the chapter, we present a phonological model of Uyghur intonation that incorporates these findings.

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 chapter 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, the responses fall into three categories based on the type of focus involved.

- **Neutral focus declaratives**:

  vidéoDA néME BOLdi? ‘What happened in the video?’

  MERyem Alimgha MÉwe BERdi ‘Meryem gave fruit to Alim’

- **Narrow, informational focus declaratives**:

  kim Alimgha MÉwe BERdi? ‘Who gave fruit to Alim?’

  MERyem Alimgha MÉwe BERdi ‘Meryem gave fruit to Alim’

- **Narrow, corrective focus declaratives**:

  MAhire Alimgha MÉwe BERdimu? ‘Did Mahire give fruit to Alim?’

  yaq, <u>MERyem</u> Alimgha MÉwe BERdi ‘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 chapter. These come from several different sources, including non-elicited conversation between pairs of Uyghur speakers during the task above, and radio broadcasts.

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 has three distinct levels of prosodic constituency: the accentual phrase (AP), the intermediate phrase (ip), and the intonational phrase (IP). A schematized representation of the hierarchical structure of these constituents and tonal markings are shown in Fig. 6.6, and the next four sections describe them in greater detail.

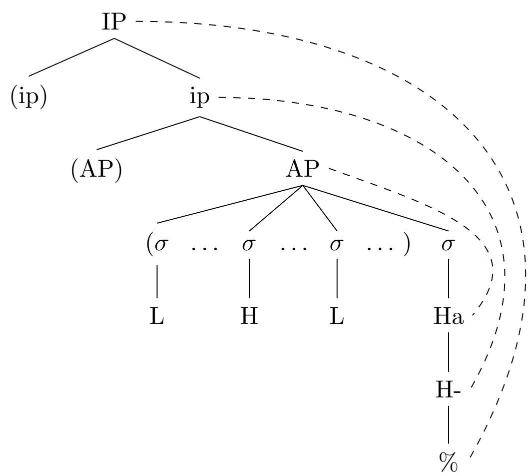

Figure 6.6: A schematic representation of the proposed prosodic hierarchy for Uyghur. Links between prosodic constituents and boundary tones are indicated by dotted lines. Prosodic tones associated with higher prosodic constituents override tones associated with lower ones when realized on the same syllable (i.e. % $\gg$ H- $\gg$ Ha).

6.4.3 The accentual phrase

The accentual phrase (AP) is the lowest level of prosodic constituency to which Uyghur intonational phonology is sensitive. An AP minimally consists of a single prosodic word, but may contain multiple words. Our data suggest that APs are generally no longer than two prosodic words.

The underlying tone pattern associated with an AP is LHLHa, where L is a low tone, H is an AP-internal high tone, and Ha is an AP-final high tone (this is similar to the tone pattern described for APs in Korean; Jun, 1998, 2000). H and Ha may be distinguished in several ways: all else being equal, the f0 peak corresponding to an Ha tone is generally higher than one corresponding to an H tone. As well, Ha tones invariably occur at the right edge of APs, while AP-internal H tones may occur elsewhere (for example, in Fig. 6.9, where an H tone occurs on the middle syllable in the verb form *MAngidu* ‘will walk’. See also Figs. 6.10, 6.15, 6.21).

The final Ha tone is invariably realized on the final syllable of the AP, except when it is overriden by a tone associated with a higher prosodic level (see Fig. 6.6). The realization of the other tones is variable, and depends partially on the number of syllables in the AP. The most common realizations in our data are LHa and LHLHa. We occasionally see other realizations such as LLHa, LHHa, Ha, HHa, and HLHa. The final three are particularly uncommon, as APs generally begin with an L tone.

There is a difference in the realization of AP tones between Uyghur and Korean, which has a similar melody for APs. In Korean, the AP-initial tone and AP-final tones are always realized. In Uyghur, however, the AP-initial tone is sometimes not realized (as in the second AP in Figs. 6.1 and 6.2 or the initial AP in Fig. 6.24), while the final tone is always realized (though it may be overwritten by boundary tones from larger constituents).

The factors determining the specific realization of AP-internal tones are not completely clear. Following Jun (2000), we transcribe the surface realizations we have encountered while leaving a careful analysis of their distribution as a topic for future research. We do note that APs containing fewer than four syllables will invariably elide one or more of the first three tones (LHL). It is common, however, for these tones to be elided even in APs with enough syllabic material to support them all. This seems to be particularly common in more rapid speech, such as the newscaster speech presented in Fig. 6.24.

Fig. 6.7 contains two APs that contain multiple prosodic words: *NÉRwa MOmay* “nervous grandmother” and *ram ONGlidi* “fixed a frame”. 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 tone, followed by an f0 descent across the verb resulting from the L% tone associated with declarative utterances (see Section 6.4.6).

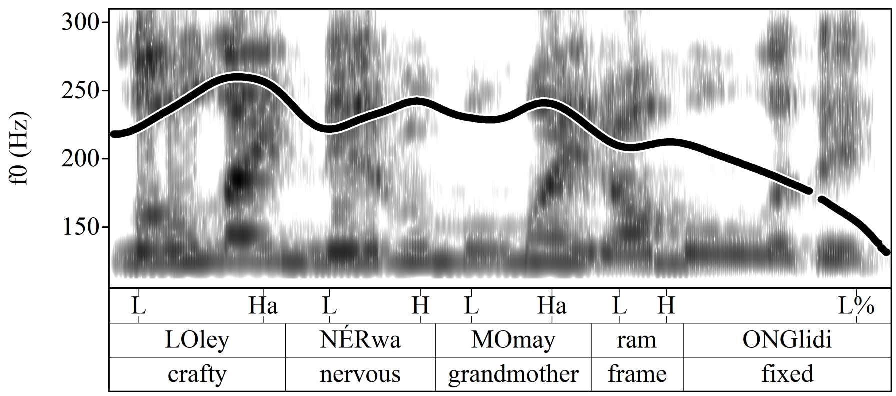

Figure 6.7: Pitch track, spectrogram, and annotation for the sentence *LOley NÉRwa MOmay ram ONGlidi* "the crafty, nervous grandmother fixed the frame."

In addition to the tonal patterns described above serving as evidence of prosodic boundaries, there are segmental phonological processes that can serve as diagnostics of these boundaries as well. One of these processes is *vowel reduction*: broadly speaking this process requires that the vowels *a* (IPA: /ɑ/) and *e* (IPA: /æ/) raise to *i* in medial, open syllables.[5]

Vowel reduction occurs most consistently at the level of the word. Examples of vowel reduction within words are shown in Tables 6.2 and 6.3.

bala	'child'	bal**i**-lar	'child-PL'
apa	'mom'	ap**i**-si	'mom-3.SG.POS'

Table 6.2: Examples of *a* vowel reduction.

apet	'disaster'	ap**i**t-i	'disaster-3.SG.POS'
méwe	'fruit'	méw**i**-ler	'fruit-PL'

Table 6.3: Examples of *e* vowel reduction.

[5]Uyghur has a second phonological process, called *umlaut* that raises the same vowels in certain contexts to [e] rather than [i]. The term *vowel reduction* is used to refer to the specific raising process discussed here.

Vowel reduction can also occur in word-final syllables: for example, in utterances like *Adil Hesenge berdi* 'Adil gave it to Hesen', the dative suffix *-ge* may raise, producing *Adil Hensengi berdi* (Hahn, 1991b). Word-medial raising is represented orthographically, while word-final raising is not.

Our data suggests that the domain for raising is the AP: that is, vowel reduction may occur at word boundaries within an AP, but not at AP boundaries. Thus vowel reduction can be used as a diagnostic tool for prosodic grouping, although there are interactions with stress and morphological complications that must be considered (see, e.g., Mayer, Eziz, & McCollum, accepted). We describe several examples below.

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 [ɑ] in *ram*, which cannot raise because it is in a closed syllable. In this token, the mean F1 value for the vowel in *ram* is 654 Hz, while the mean F1 for the final vowel in *NÉRwa* is 338 Hz.

Further examples of word-final raising can be seen in Figs. 6.18 and 6.19. In these figures (and the accompanying recordings) it is illustrative to compare the quality of the final *e* vowels in *KIMge* (Fig. 6.18 and *MERyemge* (Fig. 6.19), which both end APs, with the final *e* vowel in *MÉwe*, which does not. This latter vowel is produced with an increased F1 relative to the other vowels, indicating vowel reduction.[6]

A second diagnostic involves hiatus resolution by vowel deletion, which occurs between words within an AP, but does not occur across APs.[7] 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. The cases of hiatus resolution by vowel deletion that we have observed target the final vowel of the first word. We have not observed alternative resolution strategies such as coalescence.

We can thus differentiate AP boundaries from AP-internal word boundaries by looking at not

[6]In Fig. 6.18 it is produced more centralized than in Fig. 6.19: for a discussion of variability in the production of *i* in Uyghur, see (Mayer, Major, & Yakup, 2022).

[7]If hiatus resolution occurs across an AP or higher boundary, it is typically achieved by glottal stop insertion.

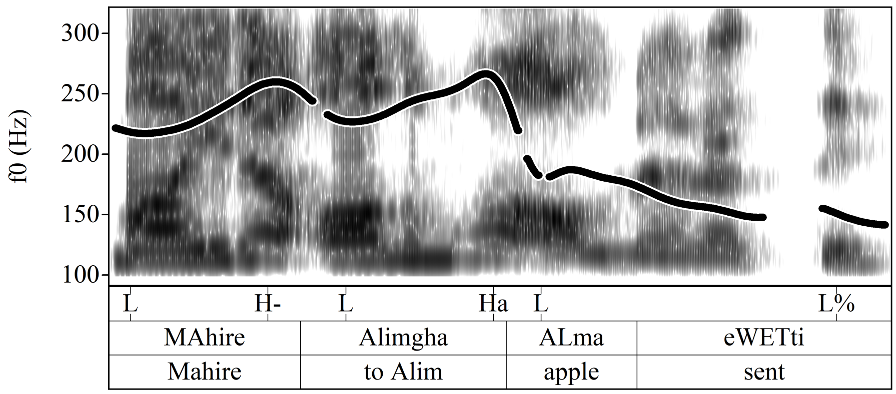

Figure 6.8: Pitch track, spectrogram, and annotation for the sentence *MAhire Alimgha ALma eWETti* "Mahire sent Alim an apple."

only the height of the H peak at the end of the phrase, but also by whether vowel reduction or hiatus resolution by deletion takes place between two words.

Finally, AP boundaries may be distinguished from AP-internal word boundaries by their greater juncture (that is, greater phrase final lengthening). The next section presents a small phonetic study that quantifies the juncture differences between these two boundary types.

6.4.4 Quantifying differences in juncture between word and AP boundaries

To provide more systematic empirical support for the claim that AP boundaries have greater juncture size than word boundaries within an AP, we carried out a small phonetic study that compares simple sentences like *qurBAN MAngidu* "Qurban will walk" (Fig. 6.9) against more complex sentences like *qurBAN biLEN ziBA MANGidu* "Ziba will walk with Qurban" (Fig. 6.10).

Notice that each of the sentences introduced above begins with the name *qurBAN*, which is the standalone subject in the first case and the complement to the postposition *biLEN* "with" in the comitative construction. Our intention is to determine whether there is a correlation between syntactic structure and the intonational properties of the first constituent.

In 6.9, *qurBAN* forms an AP. Compare this to the comitative construction in 6.10. Notice that

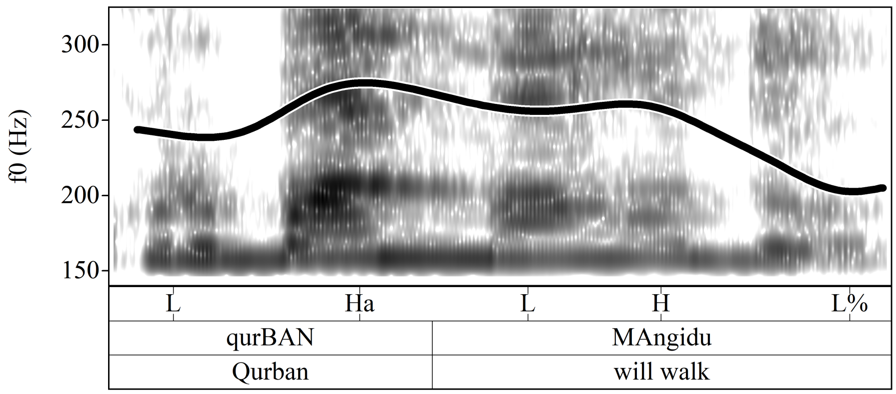

Figure 6.9: Pitch track, spectrogram, and annotation for the sentence *qurBAN MAngidu* “Qurban will walk.”

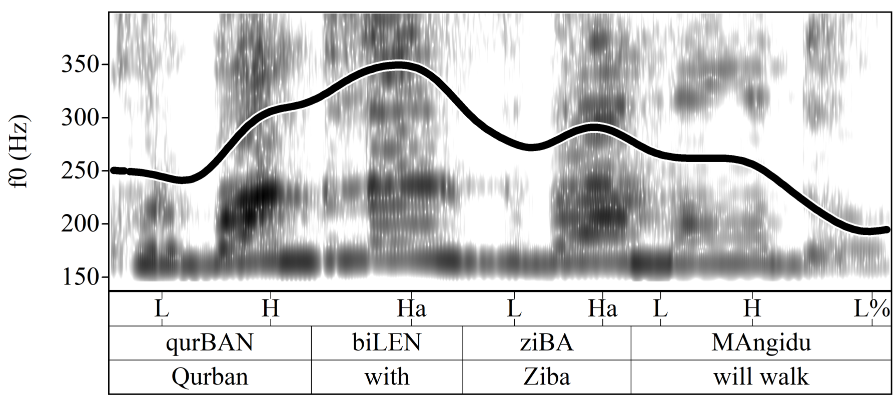

Figure 6.10: Pitch track, spectrogram, and annotation for the sentence *qurBAN biLEN ziBA MAngidu* “Ziba will walk with Qurban.”

qurBAN does not display an AP boundary tone on its right edge; instead, the Ha tone is realized on the right edge of the entire postpositional phrase. This demonstrates that there is a relationship between syntactic structure and prosodic phrasing.

We predicted that in simple sentences when the initial word is the grammatical subject of the matrix clause, then it would always form its own AP. In complex sentences when the first name is not the subject, but rather is embedded in a larger constituent such as a postpositional phrase, we predicted that the entire constituent would form an AP, with the initial word forming one of the prosodic words within the AP. These predicted differences in phrasing mean that in simple sentences the final syllable of the initial word may display greater f0 and duration due to its occurrence at the right edge of an AP than in complex sentences where it occurs AP-medially. We designed a simple experiment to test for these predicted differences.

We asked six native Uyghur speakers to read a set of 16 sentences, for a total of 96 tokens. Sentences had the following two question and response structures:

- **Simple**:

  néME BOLdi? ‘What happened?’

  <name> <verb> ‘<name> <verb>’

- **Complex comitative**:

  néME BOLdi? ‘What happened?’

  <name>$_1$ biLEN <name>$_2$ <verb> ‘<name>$_2$ <verb> with <name>$_1$’

The values for <name> were *ziBA* or *qurBAN* and the values for <verb> were *YÜridu* “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.

The final syllable of the initial word in each of these utterances was segmented using Praat. We extracted duration in seconds, mean 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 two independent variables not present in the other models. The first was whether the syllable had a coda (as in *qurBAN*) or not (as in *ziBA*); we expected that syllables with codas should have longer overall duration. The second was the total number of syllables in the utterance; we predicted that longer utterances should have shorter syllable durations (Klatt, 1973a, 1973b, 1974). We included random intercepts for each participant in all models.

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. 6.11; $\beta = 0.047, t = 3.004, p < 0.005$). Unsurprisingly, there was also a significant positive effect of coda presence on duration ($\beta = 0.040, t = 7.046, p < 0.001$). There was also a significant positive effect of number of syllables ($\beta = 0.008, t = 2.28, p < 0.05$), indicating that the final syllable of the first word was longer in longer utterances after the previous two factors were controlled for (e.g., in the complex sentences and in sentences with longer verbs). Although this effect was in the opposite direction than expected, it was quite small, and may indicate that speakers produced longer sentences more carefully. Neither the f0 or intensity models showed significant differences based on subjecthood.

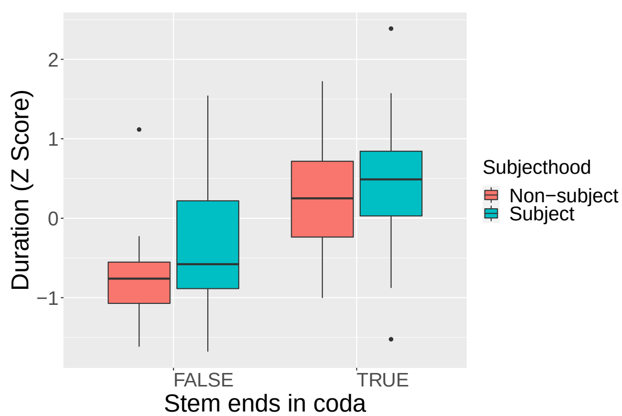

Figure 6.11: Comparison of duration of final syllables between subjects and non-subjects in word-initial position. Durations are normalized to Z scores to facilitate comparison across participants.

We interpret these differences in duration of final syllables between utterance-initial subjects and utterance-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). Here the systematic difference in syllable length between the two positions supports the claim that AP boundaries and AP-internal word boundaries may be differentiated based on juncture size.

The lack of a significant difference in f0 based on boundary type suggests that absolute f0 is

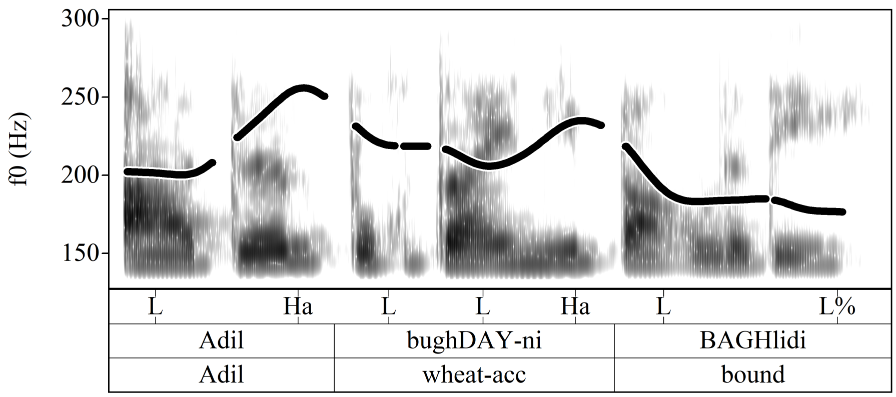

Figure 6.12: Pitch track, spectrogram, and annotation for the sentence *Adil bughDAYni BAGHlidi* "Adil bound the wheat." Note the effect of declination on the f0 peak of the second Ha tone.

not an effective means of defining Ha tones: rather the height of the peak with respect to other peaks in the same utterance must be considered.

6.4.5 The intermediate phrase

This section introduces the *intermediate phrase* (ip), the next highest level on the prosodic tree shown in Fig. 6.6. We describe the properties of the ip primarily by comparing them to the properties 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 ips and APs are that ip boundaries break declination/enable pitch range expansion and they are followed by a greater degree of juncture (similar to the difference in juncture between word boundaries and AP boundaries). With respect to pitch, H- boundaries produce an increase in the height of the following pitch peaks, rather than the H- peak itself.

Declination is a universal property of speech prosody whereby maximum f0 tends to decline over the course of an utterance (e.g. Liberman & Pierrehumbert, 1984; Prieto et al., 1996). This can be observed in many of the figures in this chapter: Fig. 6.12, for example, shows a simple example of this where the f0 peak of the first Ha tone is higher than that of the second.

ip boundaries will typically trigger pitch reset, where the trend of declination is broken and the maximum f0 of the following peak returns to roughly its utterance-initial height. An example of this is shown in Fig. 6.8, where the f0 peak at the end of *Alimgha* is roughly the same (or even slightly higher) than the peak at the edge of preceding *MAhire*.

ip boundaries are often employed to produce a pitch reset for the purpose of focusing a particular prosodic constituent (see Section 6.5.1). Their use is not limited to focus constructions, however: they may also be used to break up a long utterance into smaller phrases. Fig. 6.13 shows another instance of an ip boundary triggering pitch reset. This figure 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.[8]

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. 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. That *ziYA bazarDA Alimgha* forms a single ip is supported by both the pattern of declination within the phrase and the pitch reset observed at the beginning of the following phrase. Also, the duration of the final syllable in *Alimgha* is longer than the durations of the final syllables in *ziYA* and *bazarDA*. Because this sentence lacks a focused element, it illustrates the natural tendency to break up a long phrase (i.e., IP) into smaller phrases (i.e., ip) cued by a pitch reset.

In longer neutral contexts such as this, there seems to be considerable variability as to how the

[8] 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 <u>arrested</u> 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.

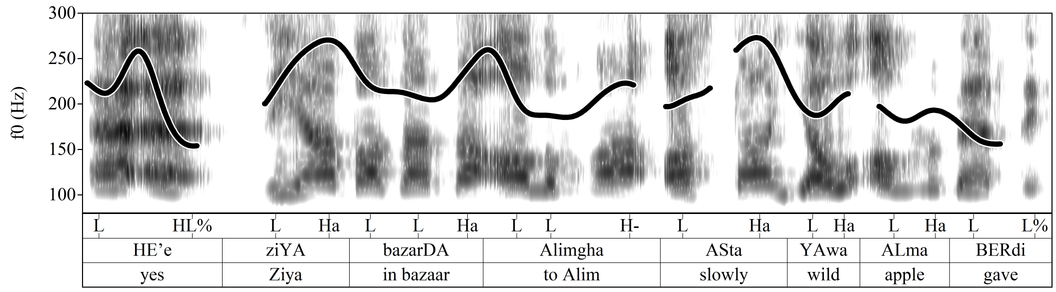

Figure 6.13: Pitch track, spectrogram, and annotation for the sentence *HE'e, ziYA bazarDA Alimgha ASta YAwa ALma BERdi* "Yes, Ziya slowly gave Alim a wild apple at the bazaar."

utterance should be broken into ips. In our investigation, we have yet to find an ip that consists of more than four 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 on focus and questions.

Finally, although there is a general tendency for ips to correspond to syntactic constituents, this is not always the case. This is also illustrated by 6.13. Despite the fact that *ziYA bazarDA Alimgha* does not form a syntactic constituent, it does form a single ip.

6.4.6 The intonational phrase

Like ips, *intonational phrases* (IP) are marked by a boundary tone on their final syllable. An IP boundary will generally cause substantial lengthening of the final syllable, be followed by an explicit pause, and trigger pitch reset. In our data, we have found four IP boundary tones: L%, H%, HL%, or LH%. The boundary tones HL% and LH% are distinct from IP-final tone sequences H L% and L H%: the complex tones are realized on the final syllable of an IP, while the tone sequences are generally realized across multiple syllables (though see additional discussion of this below). For example, Fig. 6.21 shows an LH% boundary tone while Fig. 6.23 shows an L H% sequence. Similarly, Fig. 6.22 shows an HL% boundary tone, while Fig. 6.9 shows an H L% sequence.

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, or 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.

In certain cases there is uncertainty about how tones should be analyzed when an AP consisting of a single monosyllabic word also ends an IP. Examples of this are the first word in Fig. 6.16 or the word *yoq* in Fig. 6.20. These could be analyzed as a single complex boundary tone like LH% or HL% or a sequence of two simple tones like L H% or H L%. There are cases in our data where phrase-final monosyllabic words bear multiple tones, as in the word *ram* in Figs. 6.7 and 6.15, and cases where they do not, as in *u* in Fig. 6.24. The latter case is a light syllable, however, while the other cases are heavy syllables, which may influence the number of tones that can be realized on a syllable. We interpret these ambiguous cases as sequences of simple tones for this reason.

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 prosody

Because of the absence of head-marking tones in Uyghur, focused prosodic elements are indicated by (a) positioning them at the beginning of ips such that they are preceded by a fairly large juncture and undergo pitch range expansion; (b) partial dephrasing of following material with decreased realization of optional AP-internal tones; and (c) additional pitch range expansion on the focused item and pitch range compression on the post-focus items. This is similar to the strategies used in other edge-prominence languages like Korean (Jun, 2005).[9]

Recall the simple sentence *Adil bughDAYni BAGHlidi* “Adil bound the wheat” in Fig. 6.12, which represents prototypical neutral Uyghur prosody: the first Ha on *Adil* displays the highest f0 peak in the utterance, followed by a slightly lower f0 peak for the second Ha on *bughDAYni* “wheat”, 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. 6.14. Notice that the f0 peak of the Ha tone associated with the right edge of the focused object displays substantial pitch range expansion, and the subject’s final syllable is lengthened. Both these observations suggest that the focused constituent begins an ip. Further, the juncture following (=*mu*) is smaller than that following *Adil*, suggesting that the focused constituent constitutes an AP rather than an ip.

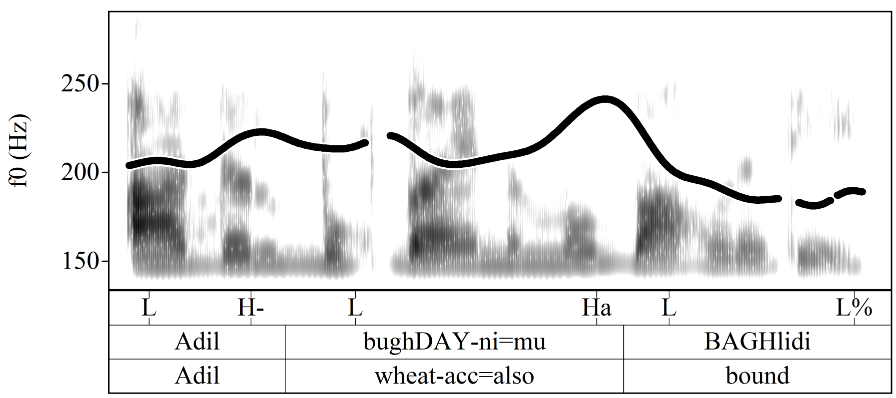

Figure 6.14: Pitch track, spectrogram, and annotation for the sentence *Adil bughdaynimu baghlidi* “Adil even bound the wheat.”

The prosody of focus when there is no explicit focus particle is essentially identical. Fig. 6.15, for example, shows a case where the adverb *büGÜN* ‘today’ receives focus. *büGÜN* is preceded by an ip boundary, which is indicated by the expanded pitch range on *büGÜN* and the phrase-final lengthening and H- boundary tone on the proceding word *MOmay* ‘grandma’.

[9]Though note again the interesting case of Farasani Arabic, which lacks head-marking tones in neutral contexts but applies pitch accents to stressed syllables in focused contexts (Abbas, 2021; Abbas & Jun, 2021).

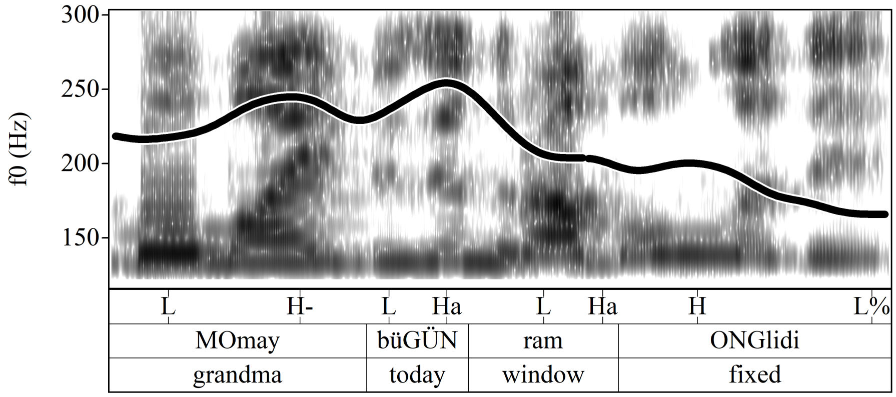

Figure 6.15: Pitch track, spectrogram, and annotation for the sentence *MOmay büGÜN ram ONGlidi* 'The grandmother fixed a frame today.'

Corrective focus is realized similarly. Consider the adverb *ASta* "slowly" in Fig. 6.16. 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. As above, the prominence of *ASta* is indicated by placing it at the beginning of a new ip. Note as well that although both *Alimgha* and *ASta* begin ips (evidenced by the preceding juncture size and pitch range expansion in both cases), the ip containing *ASta* displays a greater degree of expansion, which is typical of focused elements.

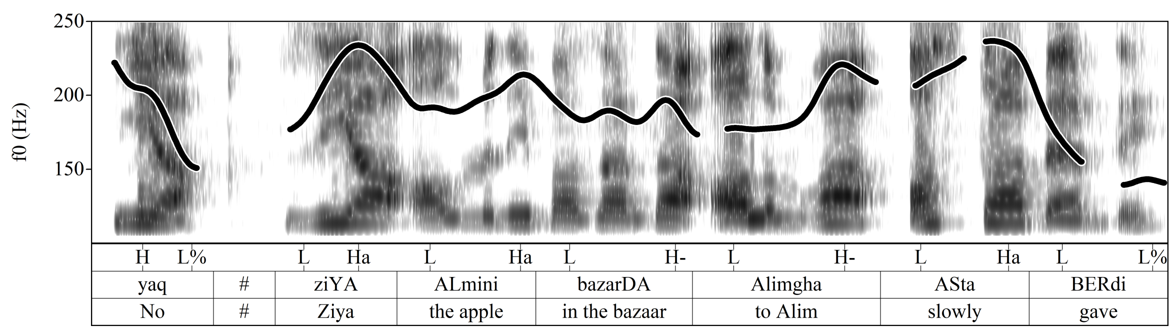

Figure 6.16: Pitch track, spectrogram, and annotation for the sentence *yaq, ziyA ALmini bazarDA Alimgha ASta BERdi* "No, Ziya slowly gave the apple to Alim at the bazaar."

6.5.2 Question prosody

Polar questions typically end in an H% or LH% tone. An example of this is provided in Fig. 6.17, which represents a simple yes/no question. In this case, the H% target is hosted by the final question particle *mu*.

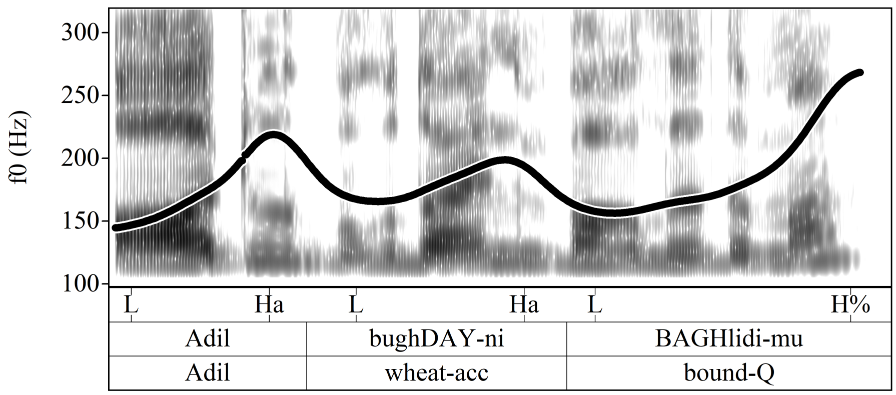

Figure 6.17: Pitch track, spectrogram, and annotation for the sentence *Adil bughDAYni BAGHlidimu?* "Did Adil bind the wheat?"

Like the polar question shown in Fig. 6.17, wh-questions are also marked with an IP-final H% or LH% tone. In addition to this final tone, wh-expressions within these questions generally receive the same kind of intonational prominence as other types of focused NPs, as demonstrated in Fig. 6.18. Here the wh-expression *KIMge* "to whom" is focused. The right edges of *Amine* and *DÜshenbe* display similar juncture sizes, while the final syllable of *KIMge*, which receives focus in this utterance, displays a lesser degree of lengthening, despite the fact that its f0 peak is the highest. This again supports the claim that focused constituents occur as the first AP in a new ip. Note as well that Fig. 6.18 demonstrates clear dephrasing and a lack of realization of several of the optional AP-internal tones following the focused element: there are no tonal targets in the AP constituting *MÉwe BERdi* aside from the initial and final tones.

The same properties are observed in the answer to Fig. 6.18, provided in Fig. 6.19. Notice that the only new information that is introduced is *MERyemge* "to Meryem", which receives

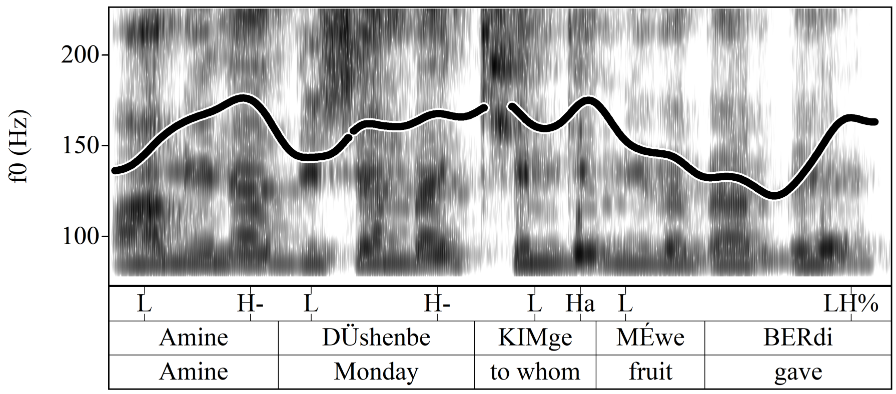

Figure 6.18: Pitch track, spectrogram, and annotation for the sentence *Amine Düshenbe KIMge MÉwe BERdi?* “Who did Amine give fruit to on Monday?”

focus. It begins the new ip following “Monday”, has the highest f0 peak in the utterance, and triggers pitch compression of the material that follows it. The only substantial difference between the question and its answer is the status of the IP-final tone, which is marked with L% in the answer because it is a declarative utterance.

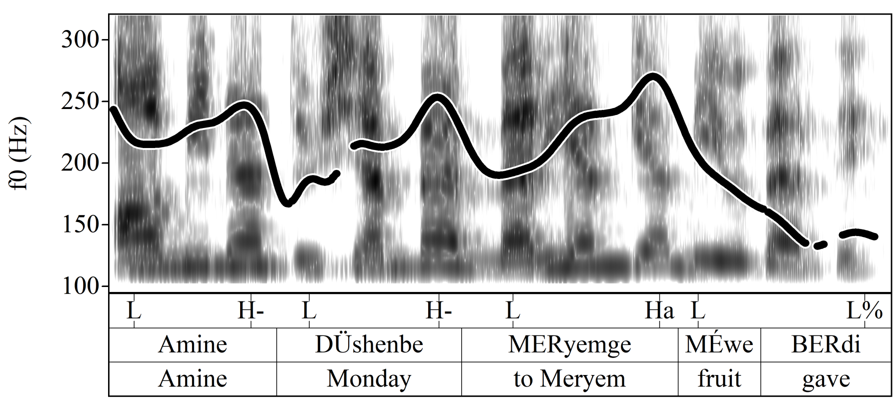

Figure 6.19: Pitch track, spectrogram, and annotation for the sentence *Amine Düshenbe MERyemge MÉwe BERdi* “Amine gave fruit to Meryem on Monday.”

6.5.3 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. 6.20 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 an 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, though "the apple" undergoes topicalization.

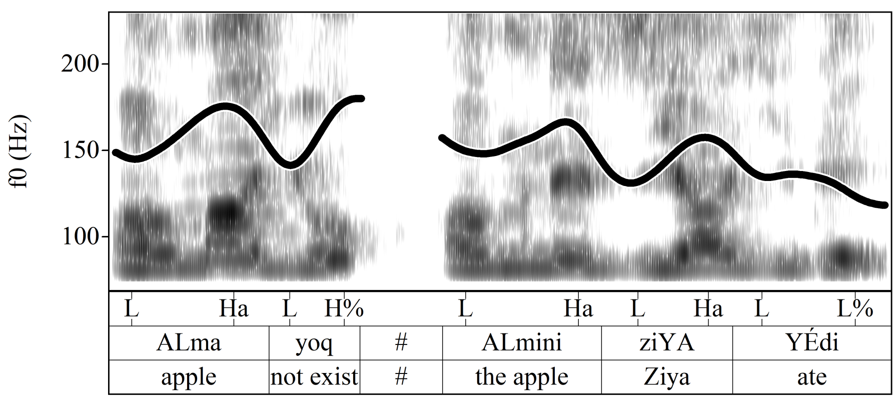

Figure 6.20: Pitch track, spectrogram, and annotation for the sentence *ALma yoq. ALmini ziYA YÉdi.* "The apple is gone. Ziya ate it."

This use of H% or LH% is also common when one gives orders or instructions; in particular, when the order or instructions consist of more than a single sentence or clause. This is exemplified by the naturalistic utterance in Fig. 6.21. 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 LH% 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).

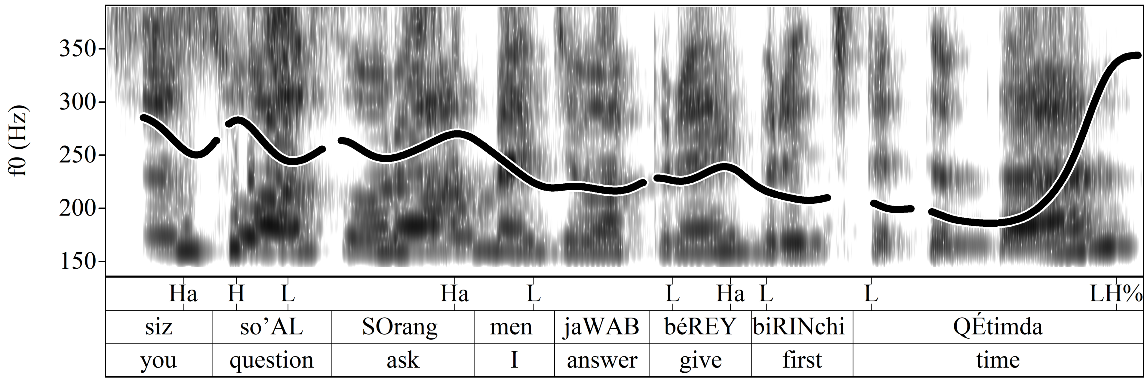

Figure 6.21: Pitch track, spectrogram, and annotation for the sentence *Siz so'AL SOrang, men jaWAB béREY, biRINchi QÉtimda* "You ask a question, I'll give an answer, for the first time." The apparent pitch trough at the end of *siz* is the result of microprosody from the [z], and the apparent peak on the [s] in *SOrang* is a pitch-tracking error.

Continuation rises are also used frequently in narratives. Consider Fig. 6.22, which shows the final two sentences of a response to a question about the speaker's experiences as a teacher.

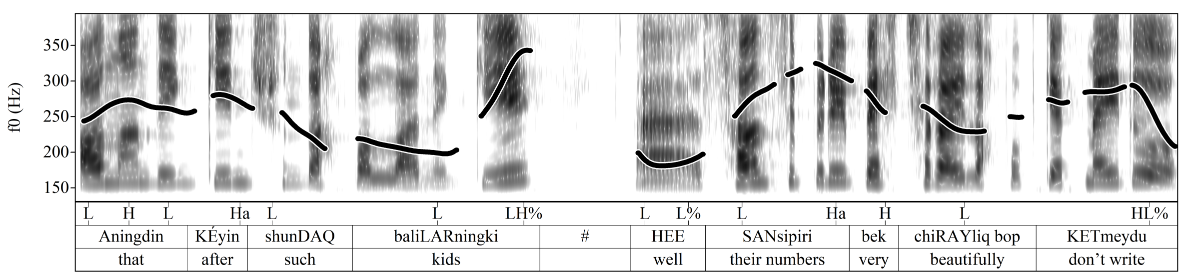

Figure 6.22: Pitch track, spectrogram, and annotation for the sentence *Aningdin KÉyin shunDAQ baliLARningki, HEE, SANsipiri bek chiRAYliq bop KETmeydu* "After that, these kinds of kids, well, they don't write their numbers very beautifully."

The first utterance 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 utterance bears the declarative HL%, indicating that she has finished speaking.

IP boundaries marked with H% or LH% are also commonly used with the converbial, clause-chaining suffix *-p* (which also surfaces as *-ip*, *-up* and *-üp* depending on its phonological context) often carries an H% or LH% tone in conjunction with a substantial juncture. This is exemplified by Fig. 6.23, 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 H% tones, reflecting the incomplete status of the utterance, while the final clause does not.[10]

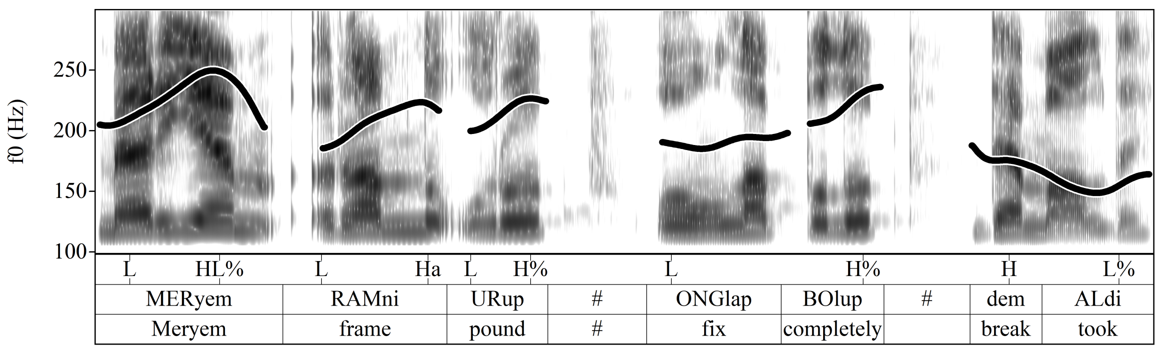

Figure 6.23: Pitch track, spectrogram, and annotation for the sentence *MERyem RAMni urUP, ongLAP bolUP, dem ALdi* "Meryem pounded the frame, fixed it completely, and then took a break."

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. It proposes that Uyghur intonation is sensitive to three prosodic levels above the word: the accentual phrase, the intermediate phrase, and the intonational phrase. Based on the previous literature, as well as the study presented in Section 6.3, the model also encodes the observation that the Uyghur intonation system is not influenced by stress, in that increased duration is the only reliable correlate of stressed syllables. This makes Uyghur unique within the set of languages that have been analyzed in the AM framework (though cf. Abbas, 2021; Abbas & Jun, 2021).

[10]The pattern in this particular instance may reflect list intonation, though similar melodies can be observed for utterances with a single instance of *-p*, where list intonation typically requires three or more elements. For discussion of the syntax of *-p* as a clause-chaining suffix, see (Major, 2021, resubmitted).

Moving forward, there are various ways in which the model can be expanded upon and improved. One crucial expansion will be the incorporation of specific break indices for annotating boundaries, leading towards a full ToBI system for Uyghur. As we have presented the model here, we have combined edge-marking tones with break indices (that is, annotating a syllable with H- tells us something about the tone as well as the size of the following juncture). Separating break indices from tones will increase the expressive power of the model, and allow more flexibility in cases where there are mismatches between tonal behavior and other properties dependant on juncture strength. It will also be important to look at a broader range of difference sentence types, such as imperatives, vocatives, echo-questions, and so on, particularly in more naturalistic contexts.

It will also be important to better understand the distribution of AP-internal tones. We claim here that the general melody for an AP is (LHL)Ha. When it is realized, the first L tone occurs on the initial syllable in the AP, and the Ha occurs on the final syllable. Whether and where the middle two tones are realized is less clear, and requires further study.

Furthermore, there are many constructions that bear interesting intonational properties and require independent investigation. There seem to be differences that vary with tense/evidentiality, agreement, clitics vs. suffixes, clausal embedding, and different types of mood marking (see, for example, Fiddler, 2021). Even in the context of focus, where we suggest that focused constituents occur at the beginning of ips, more research is necessary to determine the properties of non-focused constituents in constructions that contain focus (see, for example, the dephrasing and failure to realize optional AP-internal tones that appears to take place following the focused word in Figs. 6.1, 6.2, and 6.19).

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. Despite our model having been created and validated based on a limited set of constructions elicited specifically for this purpose, we are confident that is flexible enough as it currently stands to robustly describe the majority of intonational patterns encountered in Uyghur: that is, these outstanding issues will likely focus on predicting the prosodic realization of particular utterance types rather than substantially altering the existing structure of the model.

Take for instance the long naturalistic utterance presented in Fig. 6.24, which is from a broadcast on Radio Free Asia (Hoshur, 2010). Despite the fact that we did not explicitly discuss some of the structures present in this example, such as embedded clauses, there are no tonal patterns present that are incompatible with those discussed in this chapter (though the tone plateau on the final AP is unusual, and merits further attention).[11]

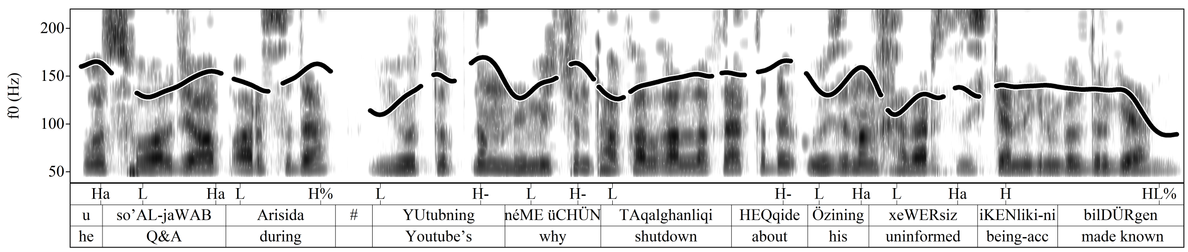

Figure 6.24: Pitch track, spectrogram, and annotation for a complex, naturalistic utterance from Radio Free Asia: *u so'AL jaWAB Arisida, YUtubning néME üCHÜN TAqalghanliqi HEQqide Özining xeWERsiz iKENlikini bilDÜRgen*, which means "During the Q&A, he let us know that he does not have any information about why YouTube was shut down."

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 elicitations 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 Uyghur speakers in Kazakhstan, a complex HL% contour at the ends of certain questions, as shown in Fig. 6.25. The general pattern found for questions involves either a gradual (H%) or sharp (LH%) rise towards the final syllable of the utterances in questions. It is unclear what function the HL% contour is serving in this case. The

[11]The initial vowel in *iKENlikini* is devoiced, which may explain why its AP does not show an initial L tone.

realization of HL% in this context also differs from the realizations in Figs. 6.22 and 6.24: in those figures the HL% boundary tone is realized entirely on the final syllable, while here it is realized across the final two syllables. Although we have analyzed such patterns as H L% sequences throughout the paper, it would be unusual for the f0 peak of the utterance to fall on an AP-internal H tone so late in the utterance.

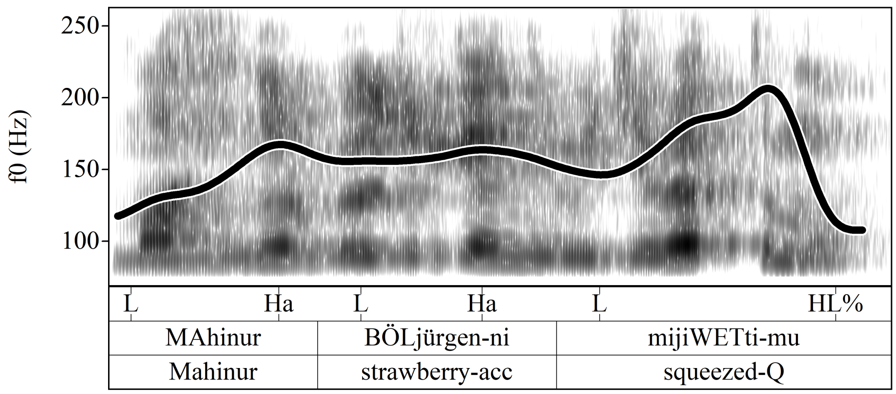

Figure 6.25: Pitch track, spectrogram, and annotation for a polar question ending in an HL% contour. This sentence means "Did Mahinur squeeze the strawberry?"

One final point worth discussing is related to the implications of the present model for future investigation of other languages. As mentioned in the introduction, Uyghur is one of few languages that has stress that is not implicated in the intonational system. For this reason, we hope that this chapter sets the groundwork that can lead to deeper investigation of other languages with similar patterns.

A summary of the tones and phrases in the model is shown in Table 6.4.

6.7 Conclusions

In this chapter, we have presented a preliminary model of the intonational phonology of Uyghur. We have presented evidence that duration is the most reliable indicator of prominence in Uyghur words, in line with Yakup (2013). Furthermore, f0 does not associate with prominent syllables,

	Phrase-internal tones
	AP-internal tones
(LHL)	Optional AP-internal tones. The first tone that is realized is typically L, and is generally aligned with the first syllable of the AP. The alignment of the other tones is variable.
	Boundary tones
	AP boundary tones
Ha	Realized on AP-final syllable
	ip boundary tones
H-	Realized on ip-final syllable
	IP boundary tones
L%	Used in declaratives; realized on IP-final syllable
H%	Used in questions and continuation rises; realized on IP-final syllable
HL%	Used in declaratives; realized on IP-final syllable
LH%	Used in questions and continuation rises; realized on IP-final syllable

Table 6.4: Inventory of tones found in Uyghur

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 accentual phrase, intermediate phrase, and intonational phrase. APs contain one or more prosodic words and have an underlying tonal melody of LHLHa, though all but the final tone may be elided depending on the size of the AP and other considerations such as speaking rate. AP boundaries also exhibit greater juncture size than AP-internal word boundaries, and prevent the application of certain phonological processes such as vowel hiatus and vowel raising. ips are distinct from APs in that their boundaries trigger pitch range expansion and have greater juncture size and lengthening of the ip-final syllable. Finally, IPs are the largest prosodic constituent in our model, whose boundary tones serve higher level functions, such as indicating 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 an earlier 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 their generosity and time in sharing their language and culture with us: Ziba Ablet, Mustafa Aksu, Akbar Akmat, Gülnar Eziz, Abduquyum Mamat, Memet Semet, Mahire Yakup, and the students of School 153 in Almaty, Kazakhstan. We would also like to thank Sun-Ah Jun, Sameer ud Dowla Khan, the attendees of Speech Prosody 9 and the Intonational Phonology of Typologically Rare or Understudied Languages, the attendees of the UCLA Phonetics Seminar, the UCSB Phonetics Circle, and two anonymous reviewers for their invaluable feedback. The authors of this chapter are listed in alphabetical order.

References

Abbas, A. (2021). *Farasani Arabic intonational phonology: A preliminary model.* (ms., UCLA)

Abbas, A., & Jun, S.-A. (2021). Prosodic structure of Farasani Arabic: Accentual Phrase without a pitch accent. In *Proceedings of the First International Conference on Tone and Intonation (TAI)* (pp. 224–228). doi: 10.21437/TAI.2021-46

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. (2021). *Praat: doing phonetics by computer [Computer Program]. Version 6.1.56.* `http://www.praat.org/`. (Accessed: 2021-11-06)

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.

Fiddler, M. (2021). The grammaticalization cline of the Uyghur converb: Syntax, morphology, and prosody. In Öner Özçelik & A. Kent (Eds.), *Proceedings of the 4th Conference on Central Asian Languages and Linguistics (ConCALL-4)* (pp. 109–118). Bloomington, IN: Center for Languages of the Central Asian Region.

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. (1998). The Accentual Phrase in the Korean prosodic hierarchy. *Phonology*, *15*(2), 189–226.

Jun, S.-A. (2000). K-ToBI (Korean ToBI) labelling conventions: Version 3. *Speech Sciences*, *7*, 143–169.

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. (1973a). Durational characteristics of prestressed word-initial consonant clusters in English. *MIT Research Laboratory of Electronics, Quarterly Progress Report*, *108*, 253–260.

Klatt, D. (1973b). Interaction between two factors that influence vowel duration. *The Journal of the Acoustical Society of America*, *54*, 1102.

Klatt, D. (1974). The duration of [s] in English words. *Journal of Speech, Language and Hearing Research*, *17*(1), 51.

Klatt, D. (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.

Liberman, M., & Pierrehumbert, J. (1984). Intonational invariance under changes in pitch range and length. In M. Aronoff & R. Oehrle (Eds.), *Language sound structure* (pp. 157–233). Cambridge, MA: MIT Press.

Lindström, E., & Remijsen, B. (2005). Aspects of the prosody of Kuot, a language where intonation ignores stress. *Linguistics*, *43(4)*, 839-870.

Major, T. (2021). *On the nature of "say" complementation* (Unpublished doctoral dissertation). University of California, Los Angeles.

Major, T. (resubmitted). Re-analyzing "say" complementation: implications for case theory and beyond. *Natural Language and Linguistic Theory*.

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 speecy 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. (2021). *Issues in Uyghur backness harmony: Corpus, experimental, and computational studies* (Unpublished doctoral dissertation). University of California, Los Angeles.

Mayer, C., Eziz, G., & McCollum, A. (accepted). Issues in Uyghur phonology. *Language and Linguistics Compass*.

Mayer, C., Major, T., & Yakup, M. (2022). Are neutral roots in Uyghur really neutral? evaluating a covert phonemic contrast. In P. Jurgec et al. (Eds.), *Supplemental Proceedings of the 2021 Annual Meeting on Phonology*. Washington, DC: Linguistic Society of America.

Menges, K. (1995). *The Turkic languages and peoples* (2nd ed.). Harassowitz Verlag.

Nadzhip, E. (1971). *Modern Uigur*. Moscow: Nauka.

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.

Prieto, P., Shih, C., & Nibert, H. (1996). Pitch downtrend in Spanish. *Journal of Phonetics*, *24*(4), 445–473.

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 stess in native speakers of Uyghur and L2 learners* (Unpublished doctoral dissertation). University of Kansas.