Achievement Predicates and Tense Paradigms in Hazaragi

Saira Bano
University of Konstanz

Miriam Butt
University of Konstanz

Ashwini Deo
Ohio State University

Proceedings of the LFG'19 Conference

Australian National University

Miriam Butt, Tracy Holloway King, Ida Toivonen (Editors)

2019

CSLI Publications

pages 394–414

http://csli-publications.stanford.edu/LFG/2019

Keywords: Hazaragi, tense/aspect, Achievements, vowel system, underspecification, morphology-syntax, lexical semantics

Bano, Saira, Butt, Miriam, & Deo, Ashwini. 2019. Achievement Predicates and Tense Paradigms in Hazaragi. In Butt, Miriam, King, Tracy Holloway, & Toivonen, Ida (Eds.), *Proceedings of the LFG'19 Conference, Australian National University*, 394–414. Stanford, CA: CSLI Publications.

Abstract

In this paper we investigate and model an interaction between lexical semantics and morphological tense paradigms in Hazaragi, an under-described Eastern Iranian language that is closely related to Dari. Hazaragi is spoken in Afghanistan and Pakistan, but also world-wide among the Hazara diaspora. Hazaragi is an SOV language with a complex tense/aspect paradigm. The paradigm exhibits a split in that it shows a past/non-past distinction with certain verbs, while with other verbs the non-past morphology is more restricted. We offer a semantic explanation for the observed pattern and show that achievement verbs lie at the heart of the split. We formally model this interaction between morphology, syntax and lexical semantics within LFG's modular architecture of grammar. In this, we rely on the proposal for the morphology-syntax interface developed by Dalrymple (2015) and the system for crosslinguistic annotation and calculation of tense/aspect developed by Zymla (2018). We implement a ParGram style grammar fragment for Hazaragi (Crouch et al. 2017, Butt et al. 1999) to model the interface between morphology, syntax and lexical semantics.

1 Introduction

Hazaragi is an Eastern Iranian language spoken mainly in the Central Afghanistan region known as Hazarajat (Dulling, 1973) and in the city of Quetta in Southern Pakistan.[1] Due to the Hazara diaspora, the language is also found across the world. Hazaragi is very close to Dari, one of the national languages of Afghanistan. It also shares features with Kabuli (Farhadi 1975), the contact vernacular in the country. While Kieffer (2003) reports Hazaragi as a dialect, *Ethnologue Languages of the World* recognizes Hazaragi as a language with a population of 2,295,000 speakers. The ethnic group is known as Hazaras or Hazareh, but the group refers to themselves as Azra and to their language as Azergi.

Hazaragi displays a complex tense/aspect system. It has two verbal stems, the present and the past stem. In this paper, we focus on a split whereby a past/non-past distinction is found in one part of the verbal paradigm, but not in another part. In example (1a), we see that the present tense form of the verb *eat* can be used to convey both ongoing present (habitual/generic or event-in-progress) temporal reference and future temporal reference. In contrast, as seen in (1b) and (1c), the past stem is restricted to past temporal reference. The imperfective can appear on both present and past stems, but it does not lead to non-past readings with the past stem. Thus, the difference in stems effects a past/non-past distinction.[2]

(1) a. ali nan mʊ-xr-a
Ali food Impf-eat.Pres-3Sg
'Ali eats food / Ali is eating food / Ali will eat food.'

[1]We gratefully acknowledge support for this research from the DAAD, the German Academic Exchange Office.

[2]The abbreviations in the glosses are as follows: Impf=Imperfective, Imp=Imperative, Subj=Subjunctive, Nom=Nominative, Pres=Present, Sg=Singular, DOM=Direct object marker.

b. ali nan mu-xord
Ali food Impf-eat.Past.3Sg
'Ali was eating food.'

c. ali nan xord
Ali food eat.Past.3Sg
'Ali ate food.'

However, the paradigm plays out differently with a certain group of verbs, as shown in (2). Here the present tense form of the verb cannot be used to convey an event-in-progress reading. It can only have habitual/generic or future reference.

(2) a. ali xana m-aj-a
Ali house Impf-come.Pres-3Sg
'Ali comes home (habitually)/#Ali is coming home/Ali will come home.'

b. ali xana amad
Ali house come.Past.3Sg
'Ali came home.'

The paradigms for the two verbs in our examples, namely *xordan* 'to eat' and *amadan* 'to come', are shown in Table 1. As can be seen, exactly the same morphology, namely the *-d*+Pers/Num affixes vs. an *mV-* prefix give rise to different tense/aspect distinctions, depending on which verb they occur with.

	xor 'eat'		*a* 'come'	
Pers/Num	**Past**	**Non-Past**	**Past**	**Non-Pres**
1st Sg.	xordʊm	mʊxrʊm	amadʊm	majʊm
2nd Sg.	xordi	mʊxri	amadi	maji
3rd Sg.	xord	mʊxra	amad	maja
1st Pl.	xordi	mʊxri	amadi	maji
2nd Pl.	xordɪn	mʊxrɪn	amadɪn	majɪn
3rd Pl.	xordən	mʊxrən	amadən	majən

Table 1: Different Tense Distinctions within the Verbal Paradigm

In section 2 we show that this split in the readings follows from an interaction of tense/aspect meanings with the Aktionsart of the verb. Achievement verbs pattern as in (2), while all other verb classes pattern as in (1). We provide a semantic explanation for this: achievements, which mostly describe punctual, non-durative eventualities, typically lack a process sub-event. When combined with present tense morphology, such verbs do not describe an eventuality in progress at reference time, since such meanings presuppose that the eventuality referred to is durative. It is this contrast between the temporal contours associated with achievement predicates vs. durative predicates (activities, accomplishments, and states)

that is responsible for the effect. While at first glance, this may appear to be a distinctly different system, one observes that achievement predicates in English inflected with present progressive marking often seem to resist a standard event-in-progress reading as well. Bach (1986) distinguishes between "culminations" and "happenings" within the class of achievements. Culminations are those predicates that describe achievements that have an associated process (*reach, die*) while happenings are 'lucky' achievements in which there is no associated process that leads to the transition described by the predicate (*notice, recognize, flash*). With culminations, progressive marking in the present tense typically induces a future-oriented reading — the event is understood to obtain in the future. With happenings, there is the added possibility of an iterative reading in which the predicate describes a sequence of repeated events. The upshot is that the peculiar temporal contour of achievement predicates is partly responsible for the unavailability of the standard event-in-progress reading associated with present marking with durative predicates. This is systematically evident in Hazaragi.

In analyzing this distinction in the verbal paradigm of Hazaragi, we were faced with complex morphophonology. We provide a first pass at an analysis of the Hazaragi verbal morphophonology in section 3.

In section 4 we show how our analysis can be implemented in a ParGram style grammar fragment for Hazaragi (Crouch et al. 2017, Butt et al. 1999). In order to model the interface between morphophonology, syntax and semantics, we rely on the proposal for the morphology-syntax interface developed by Dalrymple (2015). In section 5, we show how the tense/aspect distinctions could be calculated via Zymla's (2018) multi-tiered analysis of tense.

2 The Asymmetry in Temporal Reference

In this section we provide necessary background information on the structure of Hazaragi grammar with respect to the verbal system.

2.1 The Hazaragi Verbal Paradigm

Hazaragi is an SOV language, with both prefixes and affixes attaching to a verbal stem. The verbal paradigms are constructed on the basis of two stems: the present and the past stem. The past stem is realized via affixation of /-ɪd/ to the base form. The allomorphy in this domain is seen in Table 2. The regular verbs in lines 1–3 surface with the allomorph /-ɪd/. For irregular verbs as in the lines 4–5 the affix surfaces as /-d/ after /n/ and /r/, while after /s,ʃ,z,f/, it surfaces as /-t/ as can be seen with /ʃ/ in lines 6–7. Lines 7–8 give examples of some irregular verbs where the pattern is unpredictable.[3]

[3]There are several takes on the relationship of the stems to one another (see Farahani (1990) for an overview). One analysis postulates that the two different stems for regular verbs derive from two separate sources, another assumes that the past stem is derived from the form of the present stem via

Verb stems are further inflected for person and number. There is no gender distinction in the language. The suffixes for person and number for the past and the present are the same except for the third person singular, which has an overt inflection for the present stem, but not for the past stem (see Table 1).

	Present Stem	**Past stem**	**Gloss**
1.	bar	barɪd	pour
2.	xar	xarɪd	buy
3.	poʃ	poʃɪd	wear
4.	xor	xord	eat
5.	kan	kand	pluck
6.	koʃ	kuʃt	kill
7.	el	eʃt	keep
8.	g <gu	guft	say
9.	ʃn[4]<ʃaw	ʃud	happen

Table 2: Sample Verbs with their Present and Past Stems

There are three verbal prefixes in Hazaragi: the imperfective, the imperative/subjunctive and a marker of negation. Negation is expressed via *na*, but we are not concerned with it here. The imperfective and imperative/subjunctive surface in a variety of forms. The initial consonant is invariable: /m/ for the imperfective and /b/ for the imperative/subjunctive. This consonant is followed by a vowel which shows great variance, but this variance appears to be phonologically conditioned. We therefore analyze these two prefixes as having an underlyingly underspecified vowel: /mV-/ and /bV-/.

Imperfective marking appears on both present and past stems and gives rise to habitual/generic and progressive readings. The subjunctive and imperative are constructed by attaching the prefix /bV-/ to the present stem. While the subjunctive has forms across all the person-number cells, imperative forms are only available for the 2nd person singular and plural.

We propose that the underspecified vowel in both the imperfective /mV-/ and the imperative/subjunctive /bV-/ is subject to assimilation with the first vowel contained in the stem, thus being realized variably as [m-], [ma-], [me-], [mɪ-], [mʊ-] and [mo-]. We provide the details and motivation for this analysis in section 3, but first turn to explaining the tense split.

affixation (e.g., Lazard 1992), while another analysis (Henderson 1978) posits one underlying root which is form-identical to the non-past stem. The past stem is then taken to be derived via the affixation of an archisegment /D/, which is realized variously as /-d,-t,-ɪd/ depending on the phonological environment. The irregular verbs have no predictable pattern. Cowen and Yarmohammadi (1978) formulate three rules to derive the past stem from the underlying root.

[4]The deletion of /aw/ and insertion of /n/ indicates that several phonological processes have applied. We are currently studying these.

2.2 Explaining the Asymmetry in Temporal Reference

As already mentioned in the introduction, Hazaragi exhibits a past/non-past distinction for most verbs, as seen in (3) (repeated from above). The present tense form is ambiguous with respect to temporal reference and disambiguation can be effected via overt temporal expressions, as shown in (4).

(3) a. ali nan mʊ-xr-a
Ali food Impf-eat.Pres-3Sg
'Ali eats food / Ali is eating food / Ali will eat food.'

b. ali nan xord
Ali food eat.Past.3Sg
'Ali ate food.'

(4) a. ali darau nan mʊ-xr-a
Ali now food Impf-eat.Pres-3Sg
'Ali is eating food now.'

b. ali saba nan mʊ-xr-a
Ali tomorrow food Impf-eat.Pres-3Sg
'Ali will eat food tomorrow.'

In contrast, as shown above in (2) and repeated here in (5), with some verbs, the present tense form is not compatible with an event-in-progress reading. It only yields future-oriented or habitual/generic readings, as shown in (6).

(5) a. ali xana m-aj-a
Ali house Impf-come.Pres-3Sg
'Ali comes home (habitually)/Ali will come home/#Ali is coming home.'

b. ali xana amad
Ali house come.Past.3Sg
'Ali came home'

(6) ali ameʃa ʃau m-aj-a
Ali always night Impf-come.Pres-3Sg
'Ali always comes in the evening.'

2.2.1 Durative vs. Non-durative Verbal Predicates

A closer look at the class of verbs that fail to give rise to the event-in-progress reading with present morphology shows that it is exactly those verbs which denote achievements in the sense of Vendler (1957). Achievement verbs are taken to describe punctual events, specifically a transition to some result state. Although there is a sense in which an accompanying process is presupposed in the use of such

verbs, such a process is not taken to be lexicalized in the verb's meaning. What is critical here is that achievement predicates, in contrast to other aspectual classes, denote punctual eventualities that do not unfold over an interval of time. A useful diagnostic for identifying achievement predicates is the interaction between these and temporal expressions such as *for an hour* and *in an hour*. As has been noted starting with Dowty (1979), achievement predicates are unacceptable with *for an hour* type of temporal expressions, which presuppose atelicity. With *in an hour* type of temporal expressions, they give rise to a reading distinct from accomplishments, conveying that the eventuality described by the verb occurred *after* one hour from some salient reference time in the past. We see for Hazaragi that the achievement verb 'reach' is unacceptable with the durative *for an hour* and gives rise to the 'after an hour' reading with the *in an hour*.

(7) a. ʈrain da station məne jak genʈa rasıd
train at station in one hour reach.Past.3Sg
'The train reached the station in an hour(after an hour).'

b. ∗ʈrain da station bleje jak genʈa rasıd
train at station for one hour reach.Past.3Sg
'The train reached the station for an hour.'

Almost the same pattern of interpretation and acceptability is exhibited for *amadan* 'to come', shown in (8). The difference is that while the collocation with *for an hour* is unacceptable on the reading that the coming eventuality lasted an hour, it is acceptable on the reading that the result-state of the coming eventuality (Ali's being home) lasted for an hour. Piñon (1999) has observed that this is an available reading for durative expressions with achievement predicates. In contrast, *xordan* 'to eat' in (9) is an activity predicate (unless combined with a quantized object argument). It is acceptable with the durative *for an hour* and conveys that the eating eventuality went on for an hour. It is less acceptable with *in an hour*.

(8) a. ali məne jak genʈa xana amad
Ali.Nom in one hour house come.Past.3Sg
'Ali came home in an hour.'

b. ali bleje jak genʈa xana amad
Ali for one hour house come.Past.3Sg
'Ali came home for an hour (he stayed for an hour).'

(9) a. ∗ali məne jak genʈa nan xord
Ali in one hour nan eat.Past.3Sg
'Ali ate food in an hour.'

b. ali bleje jak genʈa nan xord
Ali for one hour nan eat.Past.3Sg
'Ali ate food for an hour.'

Other stative/activity predicates such as *ʃɪʃtan* 'to sit' and *malidan* 'to rub', which are clearly durative and atelic, pattern like *xordan* 'to eat', see (10) and (11).

(10) a. ∗ali məne jak genʈa ʃɪʃt
Ali in one hour sit.Past.3Sg
'Ali sat in an hour.'

b. ali bleje jak genʈa ʃɪʃt
Ali for one hour sit.Past.3Sg
'Ali sat for an hour.'

(11) a. ∗ali məne jak genʈa malam xo-r malɪd
Ali in one hour ointment his-DOM rub.Past.3Sg
'Ali rubbed his ointment in an hour.'

b. ali bleje jak genʈa malam xo-r malɪd
Ali for one hour ointment his-DOM rub.Past.3Sg
'Ali rubbed his ointment for an hour.'

In contrast, accomplishment predicates, illustrated in (12) by *dʒersi baftan* 'to knit a jersey' denote telic events that have both a process and a result component. As seen in (12), such a predicate is acceptable with *in an hour* but unlike with achievements, the sentence conveys that the event of knitting a jersey was accomplished within that amount of time. The combination with *for an hour* is slightly less acceptable but not ungrammatical.

(12) a. maaham məne jak genʈa dʒersi baft
Maaham in one hour jersey knit.Past.3Sg
'Maaham knitted a jersey in an hour.'

b. #maaham bleje jak genʈa dʒersi baft
Maaham for one hour jersey knit.Past.3Sg
'Maaham knitted a jersey for an hour.'

Further verbs which pattern like *amadam* 'to come' (i.e., as achievements) are: *eʃtan* 'to keep', *awurdan* 'to bring', *rasɪdan* 'to reach', *rasandan* 'to cause to reach', *xastan* 'to ask', *poʃɪdan* 'to wear', *zadan* 'to hit', *murdan* 'to die' *kuʃtan* 'to kill'. The number of simple verbs in Persian numbers in the hundreds (rather than the thousands) and we expect a similar situation in Hazaragi, given the trouble we had identifying verbal predicates that were not complex predicates.

2.2.2 Further Evidence from Complex Predicates

This fundamental pattern in the language also holds for complex predications, as illustrated in (13) and (14) for two different N-V complex predicates. In (13) with the light verb 'hit', the present event-in-progress reading is unavailable. In (14) with the light verb 'do', in contrast, the reading becomes available.

(13) Ali bini mɪ-zn-a
Ali nose Impf-hit.Pres-3Sg
'Ali will sneeze./#Ali is sneezing.'

(14) Ali darga ra taxtax mʊ-n-a
Ali door DOM knock Impf-do.Pres-3Sg
'Ali will knock at the door./Ali is knocking at the door.'

Light verbs have been shown to determine the overall Aktionsart of the complex predication (Butt 1995) and this is the case in Hazaragi as well. As illustrated in (15), when the light verb is an achievement predicate (*zadan* 'to hit'), the complex predicate is also an achievement, as shown by the interpretations associated with the temporal expressions. In (15a), we get the 'after an hour' reading. It is relevant to note here that for a class of achievement predicates, the combination with *for an hour* gives rise to the iterative reading rather than unacceptability.[5] This is the case with (15b), which conveys that there was repeated sneezing over the course of an hour.

(15) a. ali məne jak genʈa bini zad
Ali in one hour nose hit.Past.3sg
'Ali sneezed in an hour (after an hour).'

b. ali bleje jak genʈa bini zad
Ali for one hour nose hit.Past.3sg
'Ali sneezed for an hour (kept sneezing over and over).'

But when the light verb denotes an activity, e.g., *kadan* 'to do' as in (17), the complex predicate patterns with all non-achievement verbs.

(16) a. ∗ali məne jak genʈa darga ra taxtax kad
Ali in one hour door DOM knock do.Past.3sg
'Ali knocked at the door in an hour.'

b. ali bleje jak genʈa darga ra taxtax kad
Ali for one hour door DOM knock do.Past.3sg
'Ali knocked at the door for an hour.'

Having understood the reason for the different tense readings in the Hazaragi verbal paradigm, we now first turn to investigating the complex morphophonology (section 3) and then provide an implementation that models the interaction between lexical semantics, tense interpretation, and morphophonology in section 4.

[5]Such predicates denote punctual events but do not encode a transition to a result state. These are known as semelfactives (Comrie 1976, Smith 1991) or happenings (Bach 1986).

3 Hazaragi Verbal Morphophonology

The morphophonological patterns within the Hazaragi verbal paradigm are complex. Recall that the imperfective prefix can be attached to either the present or past verbal stem and that this in turn is inflected for person and number. This results in the following pattern: m+vowel+stem+suffix. The imperative and subjunctive are both expressed via a b- prefix. This b- prefix exhibits exactly the same morphophonological patterns as the imperfective prefix: b+vowel+stem+suffix. An overview of the patterns for the third person singular with the present and the past stems is provided in Tables 3 and 4.

	Past Stem	**Impf. 3rd Sg**	**Transl.**
1.	bʊrd	mʊ- bʊrd	take
2.	poʃɪd	mo-pʃɪd	wear
3.	amad	me-mad	come
4.	eʃt	me-ʃt	keep
5.	xord	mʊ- xord	eat
6.	raft	mo-raft	go

Table 3: Imperfective Forms with Past Stem for the 3rd Singular

The forms in Table 4 illustrate the variation found in the surface realization of the imperfective prefix with the past stem. This variation can be accounted for through a process in which the underspecified vowel of the prefix m+vowel acquires the place features of the stem vowel as in lines 1, 2 and 4. But in lines 3, 5 and 6, we find that the prefix holds a different vowel then the stem vowel. It is seen that the stem vowels are deleted in lines 2–4.

We further investigated the present imperfective where the variation in the prefix vowel is found at its most. Some sample verbs illustrating the variation are shown in Table 5.

	Present Stem	**Impf. 3rd Sg**	**Transl.**
1.	bər<bur	mʊ- bra	take
2.	pəʃ<poʃ	mo-pʃa	wear
3.	a <ɑj	ma-ja	come
4.	el	me-la	keep
5.	xər<xor	mʊ- xra	eat
6.	r<row	mo-ra	go

Table 4: Imperfective Forms with Present Stem for the 3rd Singular

The table illustrates several complications. One concerns the identification of the underlying vowel in the present stem. In several cases this has been lost and/or a schwa surfaces for purposes of adhering to syllabic phonotactic constraints. We therefore adduced evidence from New Persian and Middle Persian verbal stems (as well as the corresponding Hazaragi past stems) to establish the underlying forms, indicated by the < in the tables. After reconstructing the Hazaragi present stem it

is evident that the underlying vowel emerges as the prefix vowel and as the stress has to be on the prefix, the stem vowel deletes from the verbal stem as in lines 3, 4 and 6. Here resyllabification also takes place and the stem consonant becomes the coda of the first syllable as in lines 1, 2 and 5.[6]

As shown in Table 5, another complication is the patterns found with the stem vowel /a/. Instead of /a/ surfacing as the prefix vowel, several different variants in terms of mɪ-, me-, mʊ- and mo- are possible.

	Present Stem	**Impf. 3rd Sg**	**Transl.**
7.	xər<xar	me-xra	buy
8.	məl <mal	mʊ- mla	rub
9.	tərs <tars	mɪ- tərsa	scare
10.	ʃan<neʃɑn	mɪ- ʃana	sit (caus)
11.	pər <par	mo-pra	'fell'

Table 5: Stems containing an /a/ Vowel with the Imperfective

In order to understand this variation, we investigated the Hazaragi vowel system in some detail.

3.1 The Hazaragi Vowel System

The Hazaragi vowel chart in Figure 1 was constructed with original data elicited from Hazaragi speakers.[7] It constitutes the first phonological analysis of the Hazaragi vowel system.[8]

There are 8 vowels in Hazaragi i, ɪ, e, ə, a, u, ʊ, o and 24 consonants.[9] We see the vowels as being divided via a primary three-way distinction of height where all the vowels above the (dotted line) are high (i, u) and high-mid and the vowels below are low (ə, a). The most interesting are the high mid vowels /e, ɪ, ʊ, o/ as they are very close to another in height. We also see a further division into front (i, ɪ, e) and back vowels (u, ʊ, o). The /ə/ is central but at a low mid position while /a/ is low but not completely at the back or front. We therefore analyze /a/ as placeless.

We adopt the *Featurally Underspecified Lexicon* (FUL) model (Lahiri and Evers 1991, Lahiri and Reetz 2002, Scharinger et al. 2010). FUL analyzes consonants and vowels via the same place features. The place features are divided into two further sub-nodes: articulator and tongue height. This means that the height features [high] and [low] can be specified independently of the place of articulation,

[6]This is true for the bisyllabic words. For the trisyllabic words the stem vowel deletion is not seen and we are currently further investigating the stress pattern and resyllabification patterns of Hazaragi.

[7]These were average of 600 tokens of 40 monosyllabic words each repeated 5 times and elicited from three speakers (one male, two female). The vowel chart is constructed on scatter plot where the F1 (formants for tongue height) was graphed on the y axis and F2 (formants for tongue frontness) on the x axis and rotated afterwards so to give the exact location of the vowel in the oral cavity.

[8]At this point we would like to thank Aditi Lahiri for providing essential guidance in pointing us towards the potentially relevant factors in the Hazaragi vowel system.

[9]We have not added the consonant chart due to space limitations.

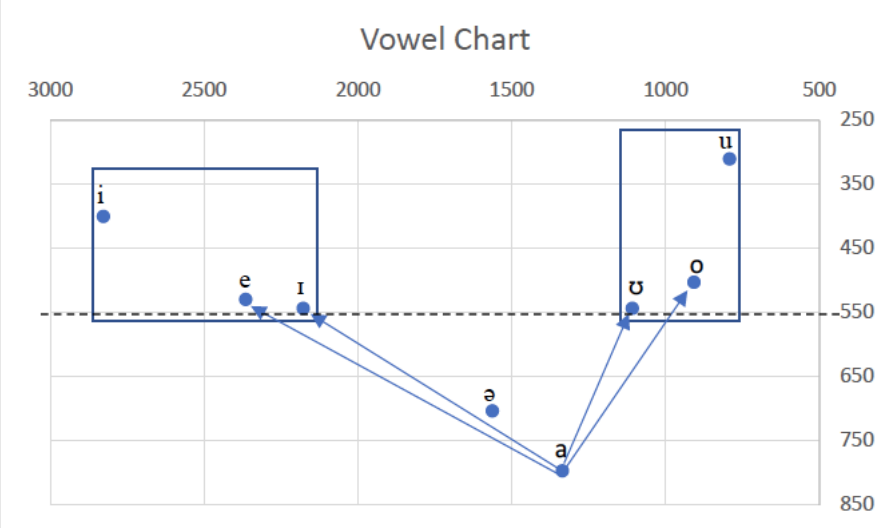

Figure 1: Hazaragi Vowel Chart

e.g. coronal, labial and dorsal. FUL also states that the height features will both be specified if the language has a three-way height distinction (Lahiri and Reetz 2010). The feature *mid* is only used for descriptive purposes. The front vowels are analyzed as coronal, the back vowels as dorsal, while rounded vowels are labial.

Applying this model to Hazaragi, we take the high-mid vowels to be underspecified for height and the low /ə, a/ vowels to be underspecified for place of articulation. Note that the high-mid vowels are the ones which emerge as the specification of the prefix vowel, which we take to be underspecified for place of articulation. Lahiri and Reetz (2002, 2010) postulate that coronal is an underspecified feature and Ghini (2001) further argues that a segment which is not specified for a place of articulation is underlyingly a coronal by default. Adopting Ghini's analysis for Hazaragi means that we assign the feature coronal to /a/ and our underspecified prefix vowel mV- by default.

An analysis based on FUL suggests separating out the surface variants of mV- into two groups of coronal vs. labial/dorsal. The [mɪ-] and [me-] are coronal, while the [mʊ-] and [mo-] are labial/dorsal vowels. The forms in Table 3 and 4 fall out naturally under this analysis: when the stem vowel is labial/dorsal, the prefix vowel is labial/dorsal. When the stem vowel is coronal, the prefix vowel is coronal. Recall that when /a/ appears as the stem vowel, the mV- surfaces variously as mɪ-, me-, mʊ- and mo-. Our analysis as the coronal being specified by default accounts for the coronal mɪ- and me- variants in Table 5.

However, something extra must be said for the mʊ- and mo- variants. With respect to these we posit that when a stem labial consonant follows the mV- it provides an environment along with the [m] of the prefix for the formation of a labial vowel in the prefix: mʊ- and mo- as in *mʊmla* in line 8 of Table 5 and the

examples in Table 6.[10]

	Present Stem	**Impf. 3rd Sg**	**Transl.**
12.	xər <xor	mʊ- xra	eat
13.	xər <xar	me-xra	buy
14.	pər <par	mo-pra	fell

Table 6: Imperfectives with Variant Forms

This initial analysis for the morphophonology of the Hazaragi verbal system suffices as a basis for this paper; In the next section, we present an implementation that models the complex interaction between lexical semantics, tense interpretation and the morphophonology of the verbal forms.

4 The Morphology-Syntax Interface

We implement our analysis via the XLE grammar development platform (Crouch et al. 2017) and integrate a finite-state morphological analyzer for Hazaragi that is based on the design and implementation of finite state morphological (FSM) analysis described in Beesley and Karttunen (2003). While we present our analysis in computational terms, we note that our treatment of Hazaragi follows the formal theoretical model for an integration of morphology into LFG as laid out by Dalrymple (2015). In that model the morphological component is assumed to specify a morphological realization relation *R*, which associates a Lexemic Index, a s(emantic)-form, and a p(honological)-form with a set of m(orphological)-features. The Lexemic index is simply an arbitrary label that is used to identify the lexeme. For our two example Hazaragi verbs *xordan* 'to eat' and *amadan* 'to come' in the imperfective present form, the morphological component relates the information in (17) to one another via the *R* relation. For the past form, the information in (18) is placed in correspondence to one another.

(17) a. M-entry for the word form *mʊxra*:
R<XOR, eats, /mʊxra/,
{M-CAT:VERB,M-STEM:PRES,M-ASP:IMPF,M-PERS:3,M-NUM:SG}

b. M-entry for the word form *maja*:
R<A, comes, /maja/,
{M-CAT:VERB,M-STEM:PRES,M-ASP:IMPF,M-PERS:3,M-NUM:SG}

(18) a. M-entry for the word form *xord*:
R<XOR, ate, /xord/,
{M-CAT:VERB,M-STEM:PAST,M-PERS:3,M-NUM:SG}

[10] /x/ is a velar in Hazaragi, but as shown that only the labial vowels or labial stem initial consonant can provide the mʊ- or mo- prefix.

b. M-entry for the word form *amad*:
$R<$A, came, /amad/,
{M-CAT:VERB,M-STEM:PAST,M-PERS:3,M-NUM:SG}

In terms of the actual FSM analysis, the finite state machine places the overt word form in correspondence with the morphological analysis as shown below. The past forms are relatively straightforward. The forms with the imperfective prefix require a more complex treatment.

```
xord:xor+Verb+Past+3P+Sg
amad:a+Verb+Past+3P+Sg
muxra:xor+Verb+Pres+Impf+3P+Sg
maja:a+Verb+Pres+Impf+3P+Sg
```

Within the FSM, the imperfective prefix is analyzed as an /m/ with an underspecified vowel: mV. The FSM specifies the list of vowels available in Hazaragi and further subdivides these into coronal (ɪ,e) and labial vowels (ʊ, o) as per the analysis in section 3. The surface realization of this underspecified vowel is calculated via a set of phonological rules that are sensitive to the first vowel and the initial consonant (labiality) found in the underlying root form of the verb.

The information from the FSM is integrated into the Hazaragi grammar via the Description function *D*, which maps a set of m-features to the appropriate c-structure category and f-descriptions. In the *dogs* example of Dalrymple (2015), this looks as in (19): For the Lexemic index DOG1 that the m-features M-CAT and M-NUM are associated with, the appropriate c-structure category is N and the f-structural information is that the number is plural. This information is passed into the grammar as part of the morphology-syntax interface.

(19) $D<$DOG1,{M-CAT:NOUN,M-NUM:PL},N,{(↑NUM)=pl}$>$

With respect to our examples, the Description function *D* relates the information shown in Table 8 to the morphological tags (m-features) coming out of the FSM. In the XLE implementation, the Description function is realized in terms of lexical entries at the sublexical level. That is, the morphological tags are treated as lexical entries with which c-structural (category) and f-structural information can be associated. For example, the entry for +Past below says that this is of the category TENSE and is associated with the f-structural information [TENSE past]. The 'xle' tells the grammar that this is information coming out of the morphological analzyer. For a more detailed description of the computational realization of this morphology-syntax interface, see Kaplan et al. (2004).

The morphological tags +Verb and +Noun do not have any f-structural information associated with them: in this analysis they simply provide category information. The +Impf tag, on the other hand, carries a complex f-structural specification. It contains a disjunction (signaled via the { | }) that says that either a future reading

or a present reading is possible, but that the present tense reading is only possible if the verb does not denote an achievement. In addition, a past tense reading is possible if licensed by the use of the past stem.

+Verb	V-T	xle	
+Noun	N-T	xle	
+1P	V-PERS	xle	(↑SUBJ PERS) = 1
+2P	V-PERS	xle	(↑SUBJ PERS) = 2
+3P	V-PERS	xle	(↑SUBJ PERS) = 3
+Sg	V-NUM	xle	(↑SUBJ NUM) = sg
+Pl	V-NUM	xle	(↑SUBJ NUM) = pl
+Past	V-STM	xle	(↑TNS-ASP TENSE) = past
+Pres	V-STM	xle	(↑TNS-ASP TENSE) ≠ past
+Impf	ASP	xle	{ (↑TNS-ASP TENSE) = pres
			(↑TNS-ASP AKTIONSART) ≠ achievement
			| (↑TNS-ASP TENSE) = fut
			| (↑TNS-ASP TENSE) =c past }

Table 7: The Description function *D*

The items specified in the sublexical lexicon need to be parsed so that only well-formed sequences of tags are allowed. This is achieved via sublexical rules that formally work just like phrase structure rules (Kaplan et al. 2004). The sublexical rule for the Hazaragi verb in our analysis is shown in Table 8. This rule expects a verb to consist of a verb stem (V-S), a verb tag (V-T), information about whether it is the present or past form of the verb stem (V-STM), an optional specification for aspect (ASP; e.g., the mV- prefix), and then person and number information.[11]

V	⟶	V-S_BASE	"verb stem"
		V-T_BASE	"category verb"
		V-STM_BASE	"past or present stem"
		(ASP_BASE)	"optional aspect (impf)"
		V-PERS_BASE	"person"
		V-NUM_BASE.	"number"

Table 8: Sublexical Rules

The verb stems themselves are specified in the lexicon as shown in Table 8. The lexical entries pick up on the lemma (Lexemic index), the category (V-S), the fact that this information is coming out of the FSM (xle) and associate the lemma

[11]The addition of _BASE is required for XLE internal reasons in the implementation to identify these rules as sublexical. We have kept them in the example so as not to sow confusion if grammar writing with XLE is attempted along the lines described in this paper. A reviewer also notes that one could in principle collapse the stem, aspect and person/number tags into one tag (V-TAG, for example) and let the FSM ensure the correct order and type of the tags. However, our experience with Urdu grammar development showed that a detailed articulation of the sublexical rules as in Table 8 did serve to constrain the grammar further.

and the category information with f-structural descriptions that include information about the predicate-argument structure and the Aktionsart of the verb.

Note that this information about the lexical semantics of a verb cannot be extracted from the morphological analyzer (since there is no morphological marking of valency or Aktionsart in Hazaragi), but should be added as part of the verbal lexicon in the grammar.

xor	V-S	xle	(↑PRED) = 'xor<(↑SUBJ)(↑OBJ)>' "eat" (↑TNS-ASP AKTIONSART) = activity.
a	V-S	xle	(↑PRED) = 'a<(↑SUBJ)(↑OBL)>' "come" (↑TNS-ASP AKTIONSART) = achievement.

Table 9: Lexical Entries for Verb Stems

With the information coming from the morphological analyzer and the information contained in the verbal stem lexicon, all the information is in place in order to provide the right analysis for the asymmetry in available readings in the Hazaragi verbal paradigm. The phrase structure rules needed for the analyses of our example sentences are simple and straightforward (and also similar to those of the Urdu ParGram grammar, see Butt and King (2007), so we do not show them here).

The f-structures in Figures 2 and 3 respectively show the analyses for the main example sentences, repeated here in (20) and (21).

(20) ali nan mʊ-xr-a
Ali food Impf-eat.Pres-3Sg
'Ali eats food/ Ali is eating food/ Ali will eat food.'

(21) ali xana m-aj-a
Ali house Impf-come.Pres-3Sg
'Ali will come home/#Ali is coming home.'

As can be seen, the combination of the imperfective prefix mV- and a non-achievement verb like 'eat' leads to ambiguity — the sentence can be interpreted either as denoting an event in the present or in the future. With an achievement verb like 'come', on the other hand, only the future reading is obtained.

This section has shown how a complex interplay between morphophonology, lexical semantics and tense readings can be modeled via a conception of the morphology-syntax interface as set out in Dalrymple (2015). For the sake of concreteness and systematic testing, we have implemented this analysis via the XLE grammar development platform, adhering to standards developed as part of the ParGram effort (Butt et al. 1999).

We have, however, so far had nothing to say about the generic/habitual readings that can still be expressed, even with achievement verbs. In the next section, we introduce Zymla's multi-tiered analysis of tense/aspect and show how this reading can be calculated on the basis of morpho-syntactic information.

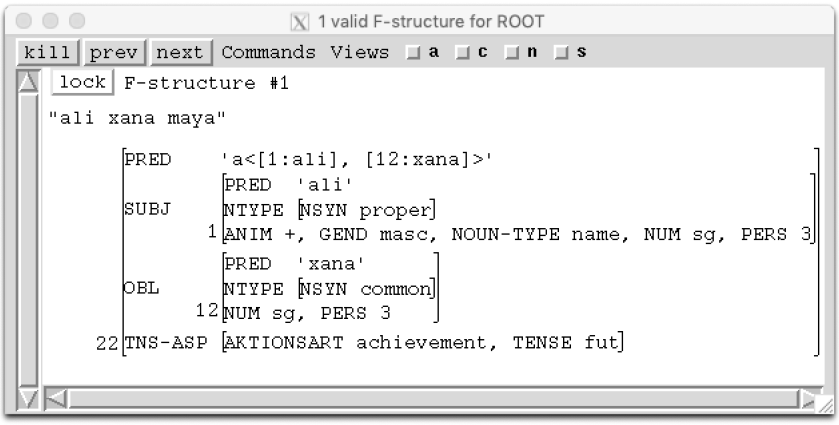

Figure 2: Present Event-in-Progress Reading Not Available

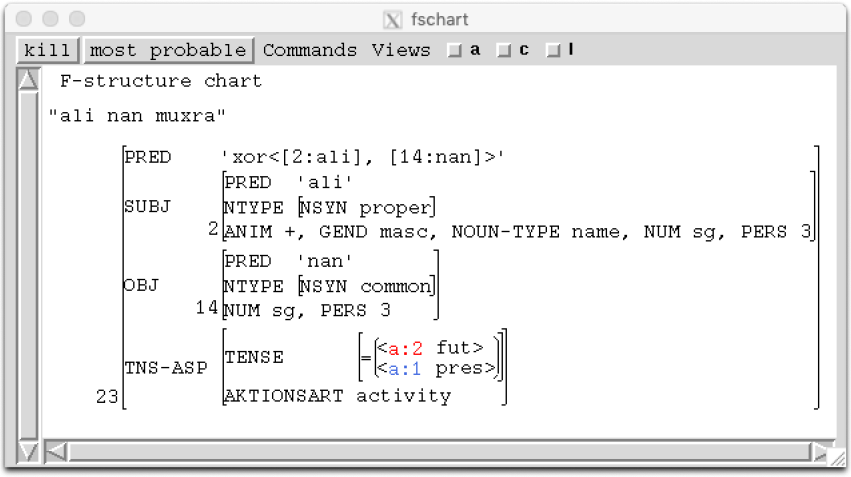

Figure 3: Present Event-in-Progress Reading Available

5 Zymla's Multi-Tiered Analysis of Tense/Aspect

5.1 The Basic System

Zymla (2018) is concerned with providing a crosslinguistically and computationally viable annotation scheme for the calculation of tense/aspect semantics. His proposal is compatible with LFG and, in particular, with ParGram style grammars (Zymla and Sulger 2017). Zymla proposes three tiers in his scheme: Tier 1 picks up on the overtly available morphosyntactic cues in a clause. For example, the verbal morphology in a sentence like *John left* indicates a simple past tense, which is registered at f-structure as [TENSE past]. On the basis of this f-structural information the semantic information can be calculated that the time of the event is to be interpreted as having occurred before the current time: $\lambda t.t \prec t_0$. A more complex

situation arises when the information relevant for a calculation of the tense/aspect semantics is derived from several different parts of the morphosyntax and is combined to derive a new meaning. An example provided by Zymla concerns habitual readings as in *John builds houses.* In this case, there is a [TENSE pres] and a plural object. From these two pieces of morphosyntactically encoded information it can be concluded that the reading is one of a habitual imperfective. This more complex calculation via a set of implication rules is accomplished at Tier 2 of Zymla's system. Tier 3 provides the locus for further pragmatic reasoning that involves putting together the information gleaned from the analysis levels provided by Tiers 1 and 2 with further information found in the clause or in the context. Zymla provides the German example in (22) as an illustration.

(22) John komm-t morgen an.
John.Nom come-Pres.3Sg tomorrow on
'John arrives tomorrow.'

As in Hazaragi, the German present morphology is in principle compatible with present event-in-progress, future and habitual/generic readings. However, unlike Hazaragi, it does not exhibit an asymmetry in readings available. In (22) the cue for disambiguation is provided by the temporal adverbial *morgen* 'tomorrow', which provides a temporal reference point that is located in the future. This part of the reasoning about the temporal information is located at Tier 3 in Zymla's system.

This multitiered system has the advantage of providing a way of calculating temporal semantic knowledge only on the basis of simple overt morphosyntactic cues (Tier 1), a combination of cues (Tier 2) and the integration of further contextual information that triggers a pragmatic reasoning component (Tier 3). If further contextual information is not available, then only reasoning up to Tier 2 is necessary. If there is no complex information to be combined, then Tier 1 calculations are sufficient. The system thus provides a nice level of "back-off" strategies depending on what type of morphosyntactic and contextual information is available.

5.2 Application to Hazaragi

In the analysis presented in section 4, the morphology-syntax interface provided f-structural tense information and already factored in the disambiguation to only future readings when the verb is an achievement (Figure 2). The analysis in section 4 had nothing to say about the habitual/generic readings that are also possible across all verb classes.

Zymla's system provides us with a slightly different way forward. Recall our central examples, repeated here in (23) and (24). An analysis at Tier 1 yields the result that we have [ASPECT imperfective] due to the mV- prefix and that we have [TENSE pres] due to the present form of the stem. A Tier 2 analysis will yield the information that the interpretation could be either in the present or in the future or be a progressive or habitual/generic predication. It will also exclude the present

and progressive interpretation for achievement verbs, as this information can be factored in by implication rules at Tier 2 that pick up on information encoded at f-structure ([AKTIONSART achievement]) via the lexical semantic specifications.

(23) ali nan mʊ-xr-a
Ali food Impf-eat.Pres-3Sg
'Ali eats food / Ali is eating food / Ali will eat food.'

(24) ali xana m-aj-a
Ali house Impf-come.Pres-3Sg
'Ali will come home/#Ali is coming home.'

Finally, calculations at the level of Tier 3 could disambiguate the possible readings by taking further contextual information such as the presence of temporal expressions as in (25) and (26) into account (repeated from above).

(25) ali ameʃa ʃau m-aj-a
Ali always night Impf-come.Pres-3Sg
'Ali always comes in the evening.'

(26) a. ali darau nan mʊ-xr-a
Ali now food Impf-eat.Pres-3Sg
'Ali is eating food now.'

b. ali saba nan mʊ-xr-a
Ali tomorrow food Impf-eat.Pres-3Sg
'Ali will eat food tomorrow.'

That is, these Tier 2 and Tier 3 calculations could be performed on the basis of the overtly available morphosyntactic information from the clause outside of the f-structural analysis. The f-structure would only represent information at the level of Tier 1 and leave the more complex calculations at Tier 2 and Tier 3 for a semantic and pragmatic component that bases itself on f-structure information, but also goes beyond it as proposed, for example, in the Abstract Knowledge Representation (AKR) approach to semantics proposed in Bobrow et al. (2007).

6 Conclusion

In this paper we have investigated the verbal paradigm of Hazaragi. We identified a split in the availability of tense readings in one and the same morphological paradigm and showed that this could be explained by taking the Aktionsart of the verbs into account. Achievement verbs lack a durative component and thus do not allow for a present tense event-in-progress reading, whereas all other verbs do.

In studying the verbal paradigm, we were confronted with complex morphophonology. The adoption of the FUL model for the Hazaragi vowel system led us

to postulate an underlyingly underspecified vowel in the imperfective prefix: mV-. We further assume that /a/ in Hazaragi is a placeless vowel and that vowels are specified as coronal by default. We argued that the underspecified vowel in the prefix acquires place features from the verbal stem vowel as well as the initial stem consonant when that is a labial. The prefix vowel thus variably surfaces as mɪ-, me-, mʊ- and mo-, as conditioned by the phonology of the stem.

With both the semantic insight and the morphophonological analysis in place, we then went on to show how the complex interaction between morphophonology, lexical semantics and tense/aspect semantics can be modeled via the morphosyntax interface as defined by Dalrymple (2015) and how tense interpretation can proceed via the multi-tier system proposed by Zymla (2018).

References

Bach, Emmon. 1986. The algebra of events. *Linguistics and Philosophy* 1, 5–16.

Beesley, Kenneth and Karttunen, Lauri. 2003. *Finite-State Morphology*. Stanford: CSLI Publications.

Bobrow, Daniel G., Cheslow, Bob, Condoravdi, Cleo, Karttunen, Lauri, King, Tracy Holloway, Nairn, Rowan, de Paiva, Valeria, Price, Charlotte and Zaenen, Annie. 2007. PARC's Bridge and Question Answering System. In Tracy Holloway King and Emily M. Bender (eds.), *Proceedings of the GEAF07 Workshop*, pages 46–66, Stanford: CSLI Publications.

Butt, Miriam. 1995. *The Structure of Complex Predicates in Urdu*. Stanford: CSLI Publications.

Butt, Miriam and King, Tracy H. 2007. Urdu in a Parallel Grammar Development Environment. In T. Takenobu and C.-R. Huang (eds.), *Language Resources and Evaluation*, volume 41, pages 191–207, Special Issue on Asian Language Processing: State of the Art Resources and Processing.

Butt, Miriam, King, Tracy Holloway, Niño, Maria-Eugenia and Segond, Frédérique. 1999. *A Grammar Writer's Cookbook*. Stanford: CSLI Publications.

Comrie, Bernard. 1976. *Aspect: An Introduction to the Study of Verbal Aspect and Related Problems*. Cambridge University Press.

Cowen, Romayne and Yarmohammadi, Lotfollah. 1978. The Persian Verb Reconsidered. *Archiv Orientalni* 46, 46–60.

Crouch, Dick, Dalrymple, Mary, Kaplan, Ronald M., King, Tracy Holloway, Maxwell III, John T. and Newman, Paula. 2017. XLE Documentation, https://ling.sprachwiss.uni-konstanz.de/pages/xle/doc/xle_toc.html.

Dalrymple, Mary. 2015. Morphology in the LFG Architecture. In Miriam Butt and Tracy H. King (eds.), *Proceedings of the LFG'15 Conference*, pages 64–83, Stanford: CSLI Publications.

Dowty, David R. 1979. *Word Meaning and Montague Grammar: The Semantics of Verbs and Times in Generative Semantics and in Montague's PTQ*. Dordrecht: Reidel.

Farahani, Ali Akbar Khomeijani. 1990. *A Syntactic and Semantic Study of the Tense and Aspect System of Modern Persian*. Ph. D.thesis, The University of Leeds.

Farhadi, A. G. 1975. *The Spoken Dari of Afghanistan: A Grammar of Kāboli Dari (Persian), Compared to the Literary Language*. Kabul.

Ghini, Mirco. 2001. Asymmetries in the phonology of Miogliola. In *Studies in Generative Grammar 60*, Berlin: Mouton de Gruyter.

Henderson, Michael M. T. 1978. Modern Persian Verb Morphology. *Journal of the American Oriental Society* 98, 375–388.

Kaplan, Ronald M., III, John T. Maxwell, King, Tracy Holloway and Crouch, Dick. 2004. Integrating Finite-state Technology with Deep LFG Grammars. In *Proceedings of the European Summer School on Logic, Language and Information (ESSLLI) Workshop on Combining Shallow and Deep Processing for NLP*.

Kieffer, Charles M. 2003. HAZĀRA iv. Hazāragi dialect. *Encyclopaedia Iranica* XII, 90–93.

Lahiri, A. and Evers, V. 1991. Palatalization and coronality. In C. Paradis and J.-F. Prunet (eds.), *The Special Status of Coronals*, pages 79—-100, San Diego: Academic Press.

Lahiri, Aditi and Reetz, Henning. 2002. Underspecified recognition. *LabPhon* 7, 637–676.

Lahiri, Aditi and Reetz, Henning. 2010. Distinctive Features: Phonological underspecification in representation and processing. *Journal of Phonetics* 38, 44–59.

Lazard, Gilbert. 1992. *A Grammar of Contemporary Persian*. Costa Mesa: Mazda Publishers.

Piñon, Chris. 1999. Durative Adverbials for Result States. In S. Bird, A. Cake, J. Haugen and P. Norquest (eds.), *WCCFL 18 Proceedings*, pages 420–433, Cascadilla Press.

Scharinger, Mathias, Lahiri, Aditi and Eulitz, Carsten. 2010. Mismatch negativity effects of alternating vowels in morphologically complex word forms. *Journal of Neurolinguistics* 23(4), 1–17.

Smith, Carlota. 1991. *The Parameter of Aspect*. Dordrecht: Kluwer Academic Publishers.

Vendler, Zeno. 1957. Verbs and Times. *The Philosophical Review* 66(2), 143–160.

Zymla, Mark-Matthias. 2018. Annotation of the Syntax/Semantics interface as a Bridge between Deep Linguistic Parsing and TimeML. In *Proceedings of the 14th Joint ACL–ISO Workshop on Interoperable Semantic Annotation*, pages 53–59, Association for Computational Linguistics.

Zymla, Mark-Matthias and Sulger, Sebastian. 2017. Cross-Linguistically Viable Treatment of Tense and Aspect in Parallel Grammar Development. In Miriam Butt and Tracy Holloway King (eds.), *Proceedings of the LFG'17 Conference*, pages 406–426, Stanford: CSLI Publications.