



Proceedings of the 2nd
Conference on Central Asian
Languages and Linguistics

(ConCALL -2)

Volume 2

7-9 October 2016
Bloomington, Indiana



Edited by
Dr. Öner Özçelik and
Amber Kennedy Kent

(C) ConCALL-2 Proceedings copyrighted by Indiana University, CeLCAR

(C) All papers copyrighted by the authors 2016

Proceedings of the 2nd Conference on Central Asian Languages and Linguistics (ConCALL-2)

Volume 2/edited by Dr. Öner Özçelik and Amber Kennedy Kent

ISBN: 978-0-9961762-1-7

1. Central Asian Languages - Congresses

For information on ConCALL-2 (held on 7-9 October 2016, Bloomington, Indiana), visit our website at
<http://www.iub.edu/~ConCALL/>

Published by

Center for Languages of the Central Asian Region

Eigenmann Hall 702

1900 E. Tenth Street

Bloomington, IN 47406-7512

USA

<http://www.iub.edu/~celcar/>

History of ConCALL

The Conference on Central Asian Languages and Linguistics (ConCALL) was founded in 2014 at Indiana University under the leadership of Dr. Öner Özçelik, the residing director of the Center for Languages of the Central Asian Region (CeLCAR), with grants and contributions from the U.S. Department of Education and several units at Indiana University, including the Ostrom Grant Programs, College of Arts and Humanities Center (CAHI), Inner Asian and Uralic National Resource, College of Arts and Sciences, School of Global and International Studies (SGIS), Sinor Research Institute for Inner Asian Studies (SRIFIAS), Department of Central Eurasian Studies (CEUS), and Department of Linguistics.

As the nation's sole U.S. Department of Education funded Language Resource Center focusing on the languages of the Central Asian Region, CeLCAR's main mission is to strengthen and improve the nation's capacity for teaching and learning Central Asian languages through teacher training, research, materials development projects, and dissemination. As part of this mission, CeLCAR has an ultimate goal to unify and fortify the Central Asian language community by facilitating networking between linguists and language educators working on Central Asian languages, encouraging research projects that will inform language instruction, and provide opportunities for professionals in the field to both showcase their work and receive feedback from their peers.

Thus, ConCALL was established to be the first international academic conference to bring together linguists and language educators working on the languages of the Central Asian region, focusing primarily on Altaic (Turkic, Mongolic, Tungusic) and Eastern Indo-European languages, among others, with the aim of encouraging research into how these specific languages are (i) represented formally, (ii) acquired by second/foreign language learners, and (iii) best taught given research driven teaching methods.

ConCALL represents all languages spoken in Central Asia and the surrounding areas, as well as languages that are genetically related to Central Asian languages. Languages represented in ConCALL-2 included, Armenian, Azerbaijani, Dagur, Dari, Farsi, Georgian, Kazakh, Kurdish, Kyrgyz, Mongolian, Pashto, Persian, Tajiki, Tibetan, Turkish, Turkmen, Uyghur, and Uzbek.

Conference Proceedings

Conference presenters were selected via an anonymous peer-review process. All conference invited speakers and accepted presenters were invited to publish their papers as part of the conference proceedings.

CONTENTS OF VOLUME 2

<i>History of ConCALL</i>	iii
<i>Contents</i>	v
<i>Editor's Introduction</i>	vii
<i>Acknowledgements</i>	viii
PART 1. PLENARY SPEAKERS	
I KAHNEMUYIPOUR, Arsalan, University of Toronto Mississauga The Ezafe Construction: Persian and Beyond	3
II MONTRUL, Silvina, University of Illinois The acquisition of “other” “different” “less common” “distant” languages: A critical need	23
III KABAK, Barış, University of Wurzburg, Germany Refining Turkish stress as a multifaceted phenomenon	25
IV ERDAL, Marcel, How did the Turkic languages come to differ as they do?	27
V KARA, György, Indiana University Lessons Learned from the Study of Mongol Languages	39
PART 2. FORMAL LINGUISTICS AND RESEARCH	
1 BOKHARI, Hassan and DURMAZ, Mustafa, Indiana University and WASHINGTON, Jonathan, Swathmore College An acoustic analysis of vowel insertion at syllable edges in Turkish	43
2 CHRISTOPHER, Nadežda, SOAS University of London Kazakh multi-functional particle ğoj: the first description	55
3 HIDAKA, Shinsuke, Tokyo University of Foreign Studies Is the -ar/-mas participle a participle in Uzbek?	69
4 KAKZHANOVA, Fazira, Karaganda State University Verbal Aspect Problems and the Aspect of the Kazakh Language	75
5 MACAK, Martin, University of Georgia The (non)realization of Armenian word-final palatal glide	85
6 YOHEI, Yamada, Tokyo University of Foreign Studies A verb aa in Dagur	95

PART 3. PEDAGOGY AND TEACHING

- 1 ANDABAYEVA, Dina and AVAKOVA, Raushangul, al-Farabi Kazakh National University 105
Language Factor In the Formation of National and Cultural Identity of Kazakhstan
- 2 ARMAN, Rahmna, KENT, Amber Kennedy and ÖZÇELİK, Öner, Indiana University 113
Project-Based Language Learning (PBL) Instruction in an Online Language Learning
Environment: A case study in Pashto
- 3 CHEN, Yunchuan and JIANG, Li, University of Hawaii at Manoa 119
The nominal expressions in Lhasa Tibetan
- 4 PUTHUVAL, Sarala, University of Washington 131
A language vitality assessment for Mongolian in China
- 5 SULEYMANOVA, Gulnara, Michigan State University 149
The acquisition of Azerbaijani idioms by second language learners

Editor's Introduction

On October 7th, 8th, and 9th, 2016, the Center for Languages of the Central Asian Region (CeLCAR) hosted the Second Conference on Central Asian Languages and Linguistics (ConCALL-2) at Indiana University in Bloomington, Indiana as the second occurrence of this biannual academic conference. ConCALL was established in May 2014 to bring together linguists and language educators specializing in the languages of the Central Asian region, including Turkic, Iranian, Mongolic, Tungusic and Tibetan languages spoken in the region, among others.

The theme of this conference, “**Continuing the Journey: Strengthening the Central Asian Language Community**,” was selected to build upon the theme of the inaugural conference which was “Building a bond: Strengthening the Central Asian Language Community” and is consistent with our goals to create a stronger network of Central Asian language experts and to provide a unique opportunity for researchers, pedagogists, linguists, and educators in the field to present their work to an audience of their peers in the same language field.

Once again, we were pleased at the overall success of the conference. We had 68 registered attendees that came from all over the globe, including not only the Central Eurasian region, including Afghanistan, Georgia, Kazakhstan, Turkey, and Uzbekistan, but also attendees from Canada, China, Germany, Japan, and the United Kingdom! Additionally, we had attendees from various American universities and institutions including Indiana University, Michigan State University, San Diego State University, University of California, Santa Cruz, University of Georgia, University of Hawaii, University of Oregon, and the University of Washington. And of course our illustrious guest speakers Marcel Erdal, Baris Kabak, Arsalan Kahnemuyipour, György Kara, and Silvina Montrul represented Goethe University, University of Würzburg, University of Toronto, Indiana University, and University of Illinois, respectively.

We were impressed and inspired by the high level of quality abstract submissions, and out of the 65 abstracts submitted, 19 were selected for oral presentations (an acceptance rate of 29.23%) and 7 were selected for poster presentations (an acceptance rate of 10.76%), with an overall acceptance rate of 40% total for both oral and poster presentations. The selected presentations and posters covered an assortment of research topics, including syntax, semantics, phonetics, phonology, language acquisition, and pedagogy, related to an assortment of language families, including Turkic, Iranian, Mongolic, Tibetan, and even Kartvelian.

We cannot express enough our gratitude to all of those who continue to support this unique conference, which we genuinely believe is making a difference to strengthen and unite our greater Central Asian language learning community, as well as increase both the quantity and quality of teaching these less commonly taught languages across the world.

And finally, we would like to once again thank the conference's founding organization, Indiana University's Center for Languages of the Central Asian Region (CeLCAR), as well as our other 2016 conference sponsors: IU's Inner Asian & Uralic National Resource Center (IAUNRC) and Center for the Study of the Middle East (CSME). Additionally, special thanks to the IU's School of Global and International Studies, College of Arts and Sciences, Central Eurasian Studies Department, and Department of Linguistics for their continued support and contributions. We look forward to seeing everyone again in Spring 2018, and hopefully some new faces as well.



Sincerely,

Dr. Öner Özçelik, Director
 Amber Kennedy Kent, Language Instructional Specialist
 Center for Languages of the Central Asian Region
 Indiana University, Bloomington, Indiana
 October 2016

ACKNOWLEDGEMENTS

The Second Conference on Central Asian Languages and Linguistics held at Indiana University on 7-9 October 2016 was made possible through the generosity of our sponsors as well as the help from the CeLCAR staff and several IU faculty and student volunteers whose diligence and hard work contributed to the successful organization and execution of the first ever ConCALL.

Major sponsors of ConCALL-2 include:

Center for Languages of the Central Asian Region (CeLCAR)

Inner Asian and Uralic National Resource Center (IAUNRC)

Center for the Study of the Middle East (CSME)

The conference would not have been as successful without their generous contributions.

Other partners we would like to acknowledge include:

School of Global and International Studies (SGIS)

College of Arts and Sciences

Department of Central Eurasian Studies (CEUS)

Department of Second Language Studies (SLS)

Center for Advanced Research on Language Acquisition (CARLA)

Additionally, we would like to thank the 25 anonymous reviewers who donated their time and expertise reviewing the conference proposals, ensuring once again the highest quality presentations possible.

Furthermore, we are eternally grateful for all of the conference presenters and participants who came from all over the globe to attend this first time conference and made it a huge success.

Finally, special thanks to all of the contributors of the ConCALL-2 Proceedings for their submissions.



Part I: Plenary Speakers

The Ezafe Construction: Persian and Beyond¹

Arsalan Kahnemuyipour
University of Toronto Mississauga

Abstract

In this paper, I explore the Persian Ezafe construction, a construction which has received significant attention in the syntactic literature in the past few decades. Descriptively, Ezafe is an unstressed vowel –e (-ye after vowels) which appears between a noun and its modifier (N-e Mod), and is repeated on subsequent modifiers, if they are present, except the last one (N-e Mod₁-e Mod₂-e Mod₃). I start with a discussion of the distribution of Ezafe, with a special emphasis on its correlation with the order of elements in the noun phrase. I provide a roll-up movement account of this construction which takes the base order of the noun phrase in Persian to be head final, with the surface order derived via phrasal movement to specifiers of intermediate functional projections. I then explore the status of Ezafe or Ezafe-like elements in several other Iranian languages to verify how this analysis fares with data from these languages.

Keywords: Ezafe, Persian, noun phrase, linker, roll-up movement

1. Introduction

There have been various accounts of the Ezafe construction in Persian. Some scholars have treated it as a case marker (Samiian, 1983, 1994; Larson and Yamakido, 2008). Other scholars have posited that it is a vowel inserted at PF to identify constituenthood or to link it to its modifiers (Ghomeshi, 1997; Ghaniabadi, 2010) or a (phrasal) affix to mark the presence of a syntactic dependent (Samvelian, 2007, 2008). There have also been accounts of Ezafe in some other Iranian languages, e.g. Zazaki (Larson and Yamakido, 2006; Toosarvandi and van Urk, 2014), Hawrami (Holmberg and Odden, 2005), Kurdish (Karimi, 2007). More recently, Ezafe has been proposed in other languages, e.g. Romanian (Irimia, Longobardi, Michelioudakis and Radkevich, 2014), Albanian (Franco, Manzini and Savoia, 2015).

In order to gain a deeper understanding of this phenomenon, in this paper, we will take close look at its realization and distribution in Persian and briefly at a few other languages. This will pave the way for exploring the “linker” phenomena in other languages in future and deciding whether all the cases labeled as Ezafe (inside and outside Iranian languages) should be treated as the same phenomenon.

2. The Distribution of Persian Ezafe

Ezafe is an unstressed vowel –e (-ye after vowels) which appears between a noun and its modifier (N-e Mod), and is repeated on subsequent modifiers, if they are present, except the last one (N-e Mod₁-e Mod₂-e Mod₃). Ezafe is present with post-nominal modifiers as shown in (1).

(1) Presence of Ezafe with post-nominal modifiers

a.	(ye)	kif-e	charm ²	b.	(ye)	mard-e	chaaq
	a	bag-Ez	leather		a	man-Ez	fat
		‘a/the leather bag’				‘a/the fat man’	

¹ The discussion of the Persian Ezafe construction in this paper is largely due to Kahnemuyipour (2014).

² Standard Persian does not have a definite article. Without the indefinite article, the unmarked reading of a modified noun phrase is definite. There is a marker of definiteness used in colloquial Persian (see Kahnemuyipour 2014).

- | | | | |
|----|------------------------------|----|-------------------------------|
| c. | sag-e qahveyi-ye gonde | d. | ketaab-e Ali/man |
| | dog-Ez brown-Ez big | | book-Ez Ali/I |
| | ‘big brown dog’ | | ‘Ali’s/my book’ |
| e. | posht-e dar | f. | Ali-ye Mohammadi ³ |
| | behind-Ez door | | Ali-Ez Mohammadi |
| | ‘behind the door’ | | ‘Ali Mohammadi’ |
| g. | gorbe-ye iraani-ye sefid-e | | Maryam |
| | cat-Ez Persian-Ez white-Ez | | Maryam |
| | ‘Maryam’s white Persian cat’ | | |

Ezafe only appears with post-nominal modifiers. It never appears on a bare noun or on pre-nominal elements. This is shown in (2) below.

(2) Absence of Ezafe with bare nouns or pre-nominal modifiers

- | | | | | | |
|----|--------------------------|----|-------------------|-------------|--------|
| a. | ketaab-(*e) ⁴ | b. | do (taa) – (*e) | ketaab | |
| | book-Ez | | two classif.-Ez | book | |
| | ‘two books’ | | | | |
| c. | in-(*e) | d. | har/hich-(*e) | ketaab-(i) | |
| | this-Ez | | each/no-Ez | book-indef. | |
| | book | | | | |
| e. | tanhaa-(*e) | f. | in-(*e) | do-(*e) | ketaab |
| | only-Ez | | this-Ez | two -Ez | book |
| | man | | | | |
| | ‘the only man’ | | ‘these two books’ | | |

When both pre-nominal and post-nominal modifiers are present, the contrast with respect to the appearance of Ezafe is evident (3).

(3) Combination of pre-nominal and post-nominal modifiers

- | | |
|----|---------------------------------------|
| a. | in do ketaab-e qatur-e jaaleb |
| | this two book-Ez thick-Ez interesting |
| | ‘these two interesting thick books’ |
| b. | in tanhaa so’aal-e bi-ma’ni |
| | this only question without-meaning |
| | ‘this only meaningless question’ |

To summarize, there is a clear correlation between the presence of Ezafe and the order of nominal elements. The noun marks a clear boundary for the Ezafe marker: all elements preceding it lack the Ezafe, while the noun itself and all elements following it (except the final one) are marked with the Ezafe. We explore this correlation in three other domains in the (morpho)syntax of Persian.

³ When speaking about a very well-known person, the Ezafe is often dropped: e.g. Ali Daayi (famous soccer player).

⁴ Note that the form in (2a) is grammatical with *-e* if *-e* is stressed and taken to be the homophonous definite colloquial marker (see Kahnemuyipour 2014). Only *-e* as Ezafe marker is relevant here.

⁵ The same adjective can be used post-nominally to mean “lonely”. The noun will have to be marked with an Ezafe, as expected. An example contrasting with (2e) is given in (ia) below. Both the pre-nominal and post-nominal adjectives can appear on the same noun, as shown in (ib).

- | | | | | |
|-----|----|------------------|----|-----------------------|
| (i) | a. | mard-e tanhaa | b. | tanhaa mard-e tanhaa |
| | | man-Ez only | | only man-Ez only |
| | | ‘the lonely man’ | | ‘the only lonely man’ |

3. Ezafe and word order: Further evidence

3.1. The superlative

While the simple/comparative forms are post-nominal and are preceded by the Ezafe marker, the superlative form is always used pre-nominally without the Ezafe vowel (4).

- (4) a. bozorg-tarin ketaab vs. ketaab-e bozorg(-tar)
 big-super. book book-Ez big-comp.
 ‘the biggest book’ ‘big(ger) book’
- b. zesht-tarin maashin vs. maashin-e zesht(-tar)
 ugly-super. car car-Ez ugly(-comp.)
 ‘the ugliest car’ ‘ugly(ier) car’

It is worth noting that examples like (5) should not be taken as counterexamples to the generalization about superlatives. In (5), the superlative is not modifying the head noun, but used in a partitive construction, with the obligatory plural marker and ambiguous singular/plural interpretation of the noun phrase even though the head noun is plural.

- (5) zesh-tarin-e maashin-*(haa)
 ugly-super.-Ez car-pl.
 ‘the ugliest of the cars’

3.2. Evidence from the formal/literary register

In the formal/literary register, adjectives can be used pre-nominally without the Ezafe vowel rather productively. No more than one adjective can be used.⁶

- (6) a. bichaare xalq vs. xalq-e bichaare
 poor/pitiable people people-Ez poor/pitiable
 ‘poor people’
- b. sabz baanu⁷ vs. baanu-ye sabz
 green lady lady-Ez green
 ‘the green lady’
- c. ... **bi-shomaar iraani-haa-yi** ke dar saraasar-e jahaan paxsh shode-and⁸
 without-number Iranian-pl.-rel. that in whole-Ez world spread become-3pl.
 ‘the innumerable Iranians who have spread all around the world’
- d. bozorg mard-e kuchak⁹
 big man-Ez little
 ‘little big man’

⁶ It appears that these pre-nominal adjectives act as a unit with the following head noun when further modified, see for example (6d). This suggests some type of compound formation.

⁷ Used in reference to Zahra Rahnavard, a prominent figure of the 2009 post-election uprising in Iran, known as the Green Movement. Zahra Rahnavard is currently under house arrest along with her husband Mir Hossein Mousavi, himself a presidential candidate of the disputed election. Another candidate of the same presidential election, Mehdi Karroubi, has also been under house arrest.

⁸ Taken from the Toronto-based Persian weekly *Shahrvand*, April 2012, issue no. 1380, page 4.

⁹ The Persian title of the 1970 Arthur Penn movie starring Dustin Hoffman.

3.3. Evidence from compound formation

In Persian compounds consisting of a modifier and a head noun, the dominant pattern is one where the adjective (or modifying noun) comes before the noun, and no *Ezafe* is used (7) (some of these are due to Ghomeshi 1996 and Lazard 1992)¹⁰:

- | | | |
|-----|--|--|
| (7) | a. gol-aab
flower-water
'rose-water' | b. bozorg-mard
big-man
'great man' |
| | c. ketaab-xune
book-house
'library' | d. kaar-xune
work-house
'factory' |

While the head-final pattern in (7) is more dominant, there are some 'compounds' with post-nominal modifiers without the *Ezafe* marker. Some such examples and their counterparts using *Ezafe* are given in (8):

- | | | | |
|-----|--|-----|--|
| (8) | a. maadar-bozorg
mother-big/grand
'grandmother' | vs. | maadar-e bozorg
mother-Ez big/grand
'big/grand mother' |
| | b. aab-porteqaal
water-orange
'orange juice' | vs. | aab-e sib
water-Ez apple
'apple juice' |
| | c. pesar-amu
son-uncle
'cousin' | | ?pesar-e amu
son-Ez uncle
'uncle's son' |
| | d. sib-zamini
apple-ground
'potato' | vs. | ?sib-e zamini ¹¹
apple-Ez ground |
| | e. tim-melli
team-national
'(Iran's) national (soccer) team' | vs. | tim-e melli
team-Ez national
'national team' |

The compounds in (8) are different from those in (7) in several ways. First, the head-final formation in (7) is much more productive than the apparently head-initial forms in (8). Second, the forms in (8) typically occur with very frequent forms only, suggesting that they originated in syntactic phrases with the *Ezafe* (N-Ez Mod) which have lost their *Ezafe* vowel over time due to frequent use (see especially (8b)).¹² Finally, the head-initial forms exemplified in (8) typically have a more transparent meaning than the head-final ones, once more underlining the syntactic phrase origin.

Given these differences, I take the head-initial compounds in (8) to be 'historical' compounds (originating as a syntactic phrase with the *Ezafe* and then losing the *Ezafe* over time), but the right-headed compounds as 'true' compounds. Crucially, I take the productive head-final compound formation of Persian as an indication for the base order of nouns and modifiers in Persian.¹³ This view is perhaps best understood in a framework that takes all word formation to occur in the syntactic component by rules of syntax – "syntax all

¹⁰ In all these cases, one could form a syntactic noun phrase involving the *Ezafe*, i.e. N-Ez A. Meanwhile, these syntactic phrases would lose the idiomatic meaning.

¹¹ The form with *Ezafe* is typically used in contrast to *sib-e deraxti* apple-Ez tree (lit. tree apple) 'apple'.

¹² In fact, with the more recent popularity of other types of juices, e.g. apple juice among many others, more head-initial forms without the *Ezafe* can be encountered. Still, if you take a fruit/vegetable, whose juice is not typically consumed, the only possible form is N-Ez N (e.g. *aab-e kaahu* juice-Ez lettuce 'lettuce juice').

¹³ Note that, in a language like English, where modifiers are consistently pre-nominal, compounds, too, are head-final.

the way down”, as held, for example, by the proponents of Distributed Morphology (Halle and Marantz, 1993; Marantz, 1997; and subsequent authors).

Note that Ghaniabadi (2010) treats the use of the post-nominal adjective (without the Ezafe marker) in the context of the colloquial definite marker (9) as a case of compounding on a par with the compounds discussed here. Kahnemuyipour (2014) argues that they should not be treated as compounds and provides an alternative analysis.

- (9) maashin zesht-e
 car ugly-def.
 ‘the ugly car’

The data illustrated so far paints the picture shown in (10) with respect to the correlation between the presence of the Ezafe and word order.

- (10) Correlation between the Ezafe and order of nominal elements
- | | | |
|------------|-----|--|
| a. N-Ez | Mod | Most common: Fully productive syntactically |
| b. Mod | N | Prenom. mod., the superlative, productive in compounds & formal register |
| c. N | Mod | Limited to ‘historical’ compounds |
| d. *Mod-Ez | N | NEVER! |

The picture in (10) shows a clear asymmetry between pre-nominal and post-nominal modifiers which requires an explanation. Extending this correlation to a more cross-linguistic perspective, we should expect the Ezafe to occur only in languages with a head-initial noun phrase and not a head-final one (in accordance with (10d)). There is some evidence in support of this view. All the West Iranian languages with Ezafe such as Persian and Kurdish dialects (e.g. Hawrami and Zazaki) have a head-initial noun phrase (see Samvelian 2008: 342). Moreover, Ezafe is not found in languages such as Germanic (e.g. English or German) or Romance (e.g. French or Italian) with head-final noun phrases.¹⁴

4. Toward an explanation of the Ezafe-word order correlation

How can we account for the (nearly) perfect correlation between the presence of Ezafe and word order both in Persian and cross-linguistically? In the generative tradition, word order variations are commonly captured via syntactic movement (displacement).¹⁵ There are two ways one can approach the current problem. One is to take the post-nominal (Persian) order as basic and derive the pre-nominal (English) order (Larson and Yamakido 2008). The other is to take the pre-nominal (English) order as basic and derive the post-nominal (Persian) order (this paper). Each of these options relates differently to the issue of the presence of overt morphology. For the first option (post-nominal base order, prenominal derived), there is complementarity between movement and overt morphology. When the Ezafe is present (Persian post-nominal order), no movement takes place; when there is no overt morphology (i.e. no Ezafe), movement derives the pre-nominal order (English). For the second option (prenominal base order, post-nominal derived), movement and overt morphology go hand in hand: When there is no movement (pre-nominal (English) order), there is also no overt morphology (no Ezafe); when movement derives the post-nominal (Persian) order, there is overt morphology (Ezafe).

¹⁴ Further investigation is needed to verify the cross-linguistic claim about the presence or absence of the Ezafe. Meanwhile, it should be noted that the claim is only about the absence of Ezafe in languages with head-final noun phrases. Therefore, the existence of languages with head-initial noun phrases and no overt realization of something akin to the Ezafe marker cannot be taken as evidence against this generalization. Of course, in Romance languages, where some modifiers can appear post-nominally, no ‘Ezafe’ is used. Also, Samvelian (2008) points to the optionality or absence of Ezafe in some (North) Western Iranian languages or Southern Kurdish dialects. The correspondence between syntactic relations and overt morphological realization is never perfect, as known in other domains such as case.

¹⁵ There is a large body of work, particularly inspired by Kayne (1994), attempting to account for this type of word order asymmetries by positing a universal base order and various (restrictions on) movement patterns (see, e.g. Cinque 1996, 1999, 2002, 2005, 2010, Belletti 2004, Rizzi 2003).

Both of the above two options have precedence in the syntactic literature: Complementarity of movement and (overt) morphology: most profoundly alluded to in Cheng's (1997) account of the typology of *wh*-questions, but also used in the realm of case. Larson and Yamakido's (2008) analysis of *Ezafe* falls into this category. According to their analysis, in Persian, the modifiers are case-marked by the *Ezafe* marker. In English, where *Ezafe* is not available, the modifier has to move to get case-licensed. We will see their analysis in more detail below.¹⁶ The view that makes a direct connection between overt morphology and overt movement is alluded to in the realm of agreement (see e.g. Kayne (1994) with respect to Hungarian adpositions, Aoun, Benmamoun and Sportiche (1994) with respect to Arabic subject-verb agreement, etc.) This is also the approach taken in this paper. Accordingly, in Persian, the overt morphology (*Ezafe*) is realized as a reflex of the overt establishment of the relation between the modifier and the noun via phrasal movement.¹⁷ When there is no overt movement (English, Persian pre-nominal modifiers), there is no overt morphology.¹⁸

4.1. *Ezafe* as case marker: Larson and Yamakido (2008)

Larson and Yamakido (2008) build on Samiiian (1994) and suggest that *Ezafe* is inserted to case-license [+N] elements, namely APs, NPs (attributive or possessive), and nominal PPs. Based on their proposal, relative clauses (CPs) and non-nominal PPs, which are not [+N], do not require case and as such, they can appear in their base position without the need for a licensing *Ezafe*.¹⁹

Some of the above discussions may already raise issues for Larson and Yamakido's analysis, but here I focus on two empirical problems, namely their treatment of relative clauses (CPs) and (non-nominal) PPs (see also Samvelian 2007, 2008). Both of these problems carry over to Samiiian (1994), who also takes *Ezafe* to be a case marker.

The first issue arises with Larson and Yamakido's (2008) treatment of relative clauses. Relative clauses, being CPs, do not require case and as such should appear in their base position without the *Ezafe* marker. They do not discuss reduced relative clauses in Persian, but based on their discussion of reduced relative clauses in English, one might expect their prediction to hold of reduced relative clauses in Persian as well. Samvelian (2007, 2008) provides counterexamples to this prediction, repeated below with minor modifications.

- (11) a. in javaan-e **[az suis bargashte]**
 this young-Ez from Switzerland returned
 'this young man who has returned from Switzerland'
- b. aks-e **[chaap-shode dar ruznaame]**
 picture-Ez published in newspaper
 'the picture published in the newspaper'

Larson (2009) argues that reduced non-finite relative clauses behave like nouns in many languages suggesting that the behavior of examples like (11) is unsurprising. Let us turn to finite restrictive relative clauses, as they are never nominal and as such predicted not to take the *Ezafe* marker. While this may appear to be true of Modern Persian, historical and typological data seem to militate against this generalization.

From a historical perspective, the Persian *Ezafe* is seen as a descendent of the Old and Middle Persian 'relative connector', used to connect the noun with the post-nominal restrictive relative clause (Samvelian 2007,

¹⁶ Larson and Yamakido (2008) do not discuss the cases of Persian pre-nominal adjectives that were considered above. If adjectives are base-generated post-nominally, it is not clear how they end up in the pre-nominal position in these cases. Given that some of these occur at the level of word formation, it may be difficult to attribute their alternative order to case.

¹⁷ The *Ezafe* has an invariant form in Persian (-e or -ye after vowels). In some other Iranian languages/dialects with richer morphology (e.g. Kurdish dialects such as Hawrami, Kurmanji, and Zazaki) the form of the *Ezafe* varies depending on phi features and other properties of the noun (see Kurmaji examples in (13) below, also Holmberg and Odden 2005, Larson and Yamakido 2006, and Samvelian 2007, 2008). These may be taken as support for the agreement view of *Ezafe*. We return to more examples below.

¹⁸ In Kahnemuyipour (2014), I propose that it is only phrasal movement and not head movement that leads to the appearance of *Ezafe*.

¹⁹ From a cross-linguistic perspective, they argue that languages such as English also have a head-initial base order for their noun phrases. Meanwhile, as they lack an overt case marker like the Persian *Ezafe*, the [+N] elements, including adjectives will have to undergo movement to get case.

²⁰ Many thanks to Azita Taleghani for bringing this fact to my attention and for an interesting discussion.

- (18) a. saalon-e [por az jam'iiyat]-e sinemaa
 hall-Ez full of population-Ez cinema
 'the movie theater filled with people'
- b. raftaar-e [xeyli dur az entezaar]-e shahrdaar
 behaviour-Ez very far from expectation-Ez mayor
 'the mayor's totally unexpected behaviour'

The existence of phrasal modifiers rules out the possibility of treating all post-nominal modifiers as heads. As we will see below, post-nominal modifiers are taken to be XPs which reside in the specifiers of functional projections above the noun. Under this view, in accordance with Bare Phrase Structure (Chomsky 1995), a bare adjective is treated as A/AP and can occupy a structural position similar to that of an AP with a complement.²⁶

Below, I develop a phrasal movement analysis of the Ezafe construction using what is known in the literature as roll-up movement (see, for example, Cinque 2005, 2010, Pearce 2002, Shlonsky 2004, etc.).²⁷

4.3. Roll-up movement analysis of the Persian Ezafe construction

Cinque (2010) develops a system in which the base order of the noun phrase is universally head final. In this system, adjectives have two possible sources: direct and indirect modification. Direct modification is lower in the structure, rigidly ordered and is associated with individual-level, nonrestrictive and nonintersective readings. Indirect modification is higher structurally, has a reduced relative clause as its source, is not rigidly ordered and has stage-level, restrictive and intersective readings. For example, in *visible visible star*, the first *visible* means 'visible now' (stage-level), whereas the second one means 'generally visible' (individual-level). An example such as *his unsuitable acts* could mean 'those of his acts which were unsuitable' (restrictive) or 'all of his acts which happened to be unsuitable' (nonrestrictive). Finally, in *the beautiful beautiful dancer*, the first occurrence of *beautiful* refers to the beauty of the person (intersective) and the second one to the beauty of the dancing (nonintersective). A slightly revised version of Cinque's general schema given in (19) is adopted here.²⁸

²⁵ While we have seen examples of nouns taking PPs (as adjuncts) requiring the Ezafe marker, some nouns with a more thematic relation with the PP seem to allow both possibilities, with or without the Ezafe. Some examples are given in (i) below. I leave the analysis of such cases for future research but it does appear that there is a subtle semantic difference between the two options.

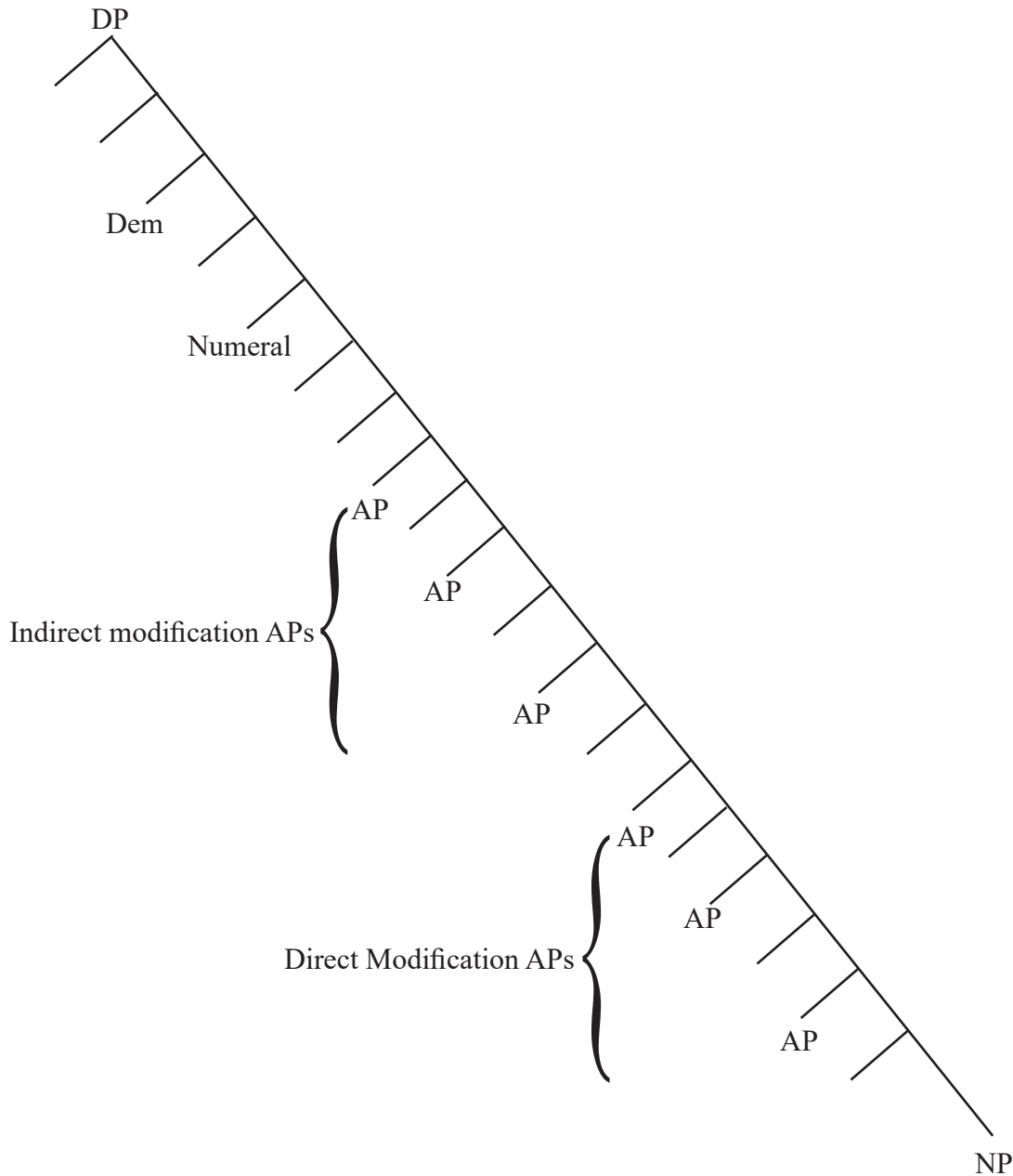
- (i) a. bahs(-e) bâ Hassan
 argument-Ez with Hassan
 b. porsesh(-e) az ra'is-jomhur
 question-Ez from president

²⁶ This uniform analysis of bare adjectives and phrasal modifiers in Persian has been challenged by Ghaniabadi (2010). For Ghaniabadi, bare adjectives are heads and head-adjoined to the noun, while AP/PP modifiers are phrasal in the specifiers of functional projections above the NP. In Kahnemuyipour (2014), I provide several arguments against his differential treatment of bare and phrasal modifiers.

²⁷ Holmberg and Odden (2005) propose a 'roll-up' derivation of the Izafe construction in Hawrami which differs in many details from the possibilities discussed in this paper. I am abstracting away from their proposal here.

²⁸ It is worth noting that whether the source of the adjective ordering is due to a universal base order (à la Cinque) or anchored in semantic notions such as scope (see, for example, Ernst 2001) is beside the point here. The crucial point for the discussion below is the existence of such an order.

(19) Structure of DP (adapted from Cinque 2010)

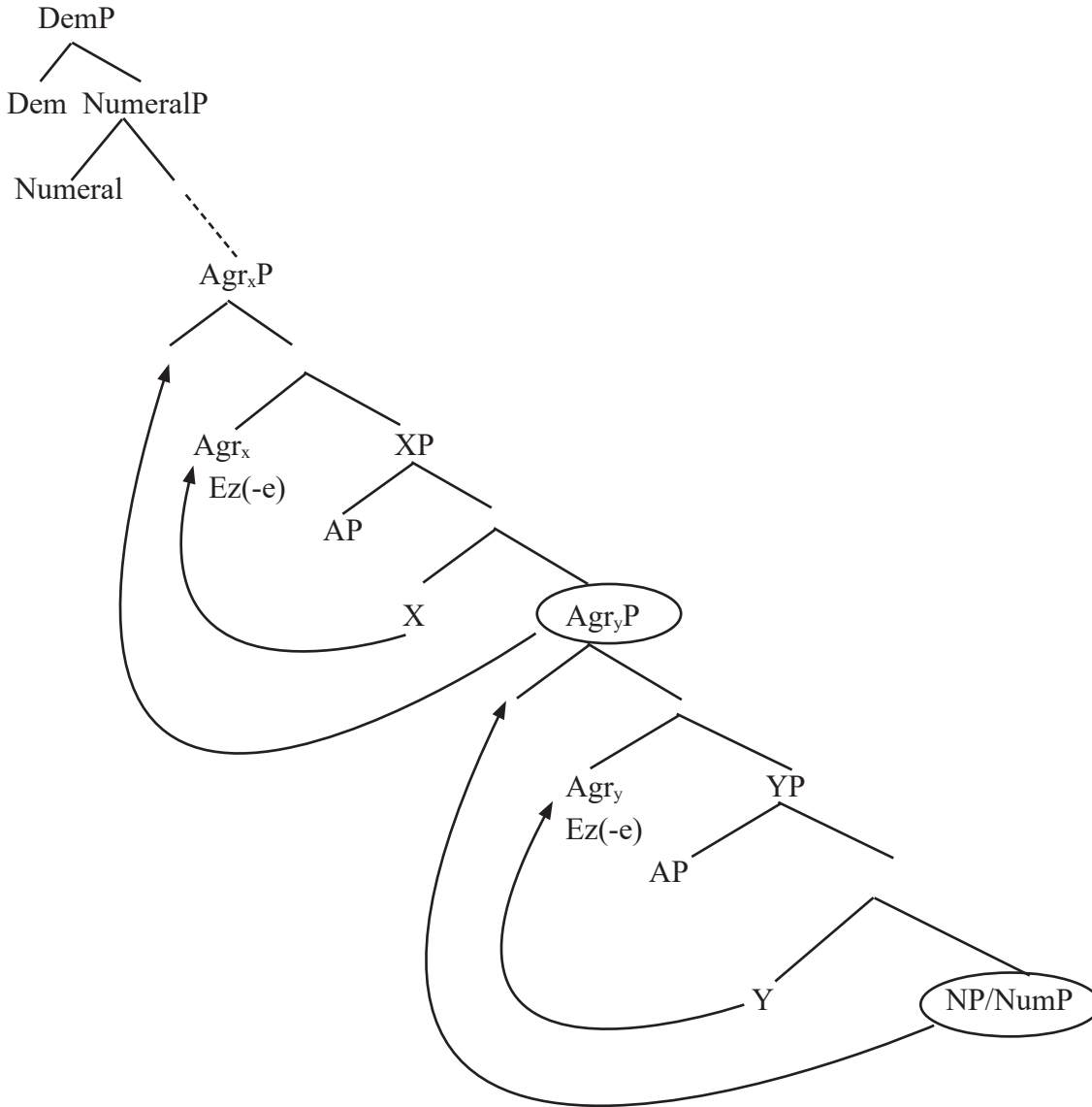


For Cinque, any variation of the order shown in (19) is the result of phrasal movement in a roll-up fashion. I extend the roll-up analysis to the Persian Ezafe construction. As such, the Persian DP is taken to be head-final, with the NP merged at the bottom of the tree structure and the APs residing in the specifiers of projections above it.²⁹ The Demonstrative (Dem) and the Numeral are heads higher up in the tree structure in accordance with (19). In addition, there are intermediate AgrP projections enabling the roll-up derivation (following Cinque 2010, among many others). As we will see below (also (13) above), in some Iranian languages/dialects, Ezafe shows morphological agreement, providing support for the Agreement status of the intermediate projections. The relevant structures and roll-up movements are shown schematically in (20).

²⁹ I am assuming that number is realized on N and moves as part of the NP complex. Note that the plural suffix is tagged along with the noun in the Ezafe construction. If one takes number (Num) to head its own projection (Ritter 1992), then the roll-up movement should start with NumP in Persian, with N raising to Num first.

Under this view, the Ezafe marker can be seen as the surface realization of the suggested inversion process, i.e. a linker in the sense of den Dikken (2006). Crucially, the height of the movement corresponds to the realization of the Ezafe marker. In this framework, the order of elements found in a language like English can be seen to reflect the base generated order of elements and as involving no movement at all (see, for example, Cinque 2010). Alternatively, one can take English to involve the same movements ‘covertly’. The Ezafe marker then finds an explanation in the context of the correlation between (overt) movement and morphology discussed above.

(20) Deriving the Ezafe construction via roll-up movement



The roll-up derivation shown in (20) combined with the base structure in (19) leads to some predictions about the order of adjectives in Persian. According to Cinque (2010), direct modification adjectives are rigidly ordered while indirect modification adjectives are not. Cinque argues that many adjectives can have either a direct or an indirect modification source. As such, while “the big red dog” is the preferred order, as it reflects the order of adjectives within the domain of direct modification, “the red big dog” is also acceptable because “red” can be used as an indirect modifier too. The same is true of Persian. Thus, both *sag-e qermez-e bozorg* (dog-Ez

red-Ez big) and *sag-e bozorg-e qermez* (dog-Ez big-Ez red) are acceptable.³⁰ Crucially, if unambiguously direct modification adjectives are used, then English shows a strict order and the roll-up analysis makes the prediction that Persian should exhibit the mirror-image order. This prediction is borne out, as shown in the examples in (21).³¹

- (21) a. mohandes-e barq-e javaan
 engineer-Ez electrical-Ez young
 *mohandes-e javaan-e barq
 ‘a young electrical engineer’
- b. mashrubxor-e sangin-e qabli
 drinker-Ez heavy-Ez former/previous
 *mashrubxor-e qabli-ye sangin³²
 ‘the former/previous heavy drinker’
- c. dalil-e asli-ye qeyre-qaabele-qabul
 reason-Ez main-Ez unacceptable
 ??dalil-e qeyre-qaabele-qabul-e asli
 ‘the unacceptable main reason’

4.4. *The missing Ezafe: Superlatives and ordinals in Persian*

Recall that the superlative, as shown in (22), is consistently pre-nominal and as such is not part of the Ezafe construction (see also Samiiian 1983, Ghomeshi 1997, Ghaniabadi 2010).

- (22) zesht-tarin maashin vs. maashin-e zesht-tar
 ugly-SUP book car-Ez ugly-COMP
 ‘the ugliest car’ ‘uglier car’

The behavior of the superlative is especially interesting given that the derivationally related simple and comparative forms of the adjective are post-nominal and require the Ezafe. This derivational relation is particularly important in the context of a non-lexicalist theory which takes word formation to be part of syntax (i.e. syntax all the way down, a la Distributed Morphology, Halle and Marantz 1993, and subsequent authors), undermining the possibility of treating the superlatives, on the one hand, and simple and comparative adjectives, on the other, as distinct lexical items which are merged in different positions of the structure (see, for example, Ghaniabadi 2010).

Cinque (2010, 31-32) discusses superlatives as one of the cases where the strict order of direct modification adjectives can be violated: a long white plane vs. the whitest long plane. More strikingly, this reversal of order is observed even with the more rigidly ordered non-intersective adjectives: an occasional hard worker vs. the hardest occasional worker. Cinque suggests that the superlative morpheme is merged high in the structure of DP and the relevant adjective is attracted to this high position (see also Matushansky 2008, among others). Extending Cinque’s analysis to Persian, I propose that in the formation of the superlative, the adjective is attracted to the superlative morpheme which is high in the DP structure. As such, the superlative adjective, like other high elements such as the demonstrative or the numeral lies outside of the Ezafe domain, i.e. it is consistently pre-nominal and lacks Ezafe.

³⁰ Some speakers report an acceptability difference between Persian and English with respect to the variability of the order of adjectives. In other words, while in English, even if both orders are possible, one is highly preferred in the unmarked case, in Persian, the preference is undermined, if not totally lost. Why should this be? It would be interesting to see if there is a general difference between languages (or adjectives) which reflect the base order and those which are derived via roll-up movement. I leave a more thorough investigation of this issue for future research.

³¹ This discussion is inspired by Cinque’s (2010) presentation of similar facts in English and Italian.

³² This is ungrammatical for the relevant sense. In the given order, “heavy” could only have a predicative meaning, i.e. heavyweight. This is true for both English and Persian.

The behavior of ordinals in Persian lends further support to the analysis of the superlative. There are two ways to express an ordinal phrase in Persian, as shown in (23).

- (23) a. *soaal-e sevvom*
 question-Ez third
 ‘the third question’
 b. *sevvom-in soaal*
 third-in question
 ‘the third question’

Once again, we see the correlation between the order of noun and modifier and the presence of Ezafe. Interestingly, the same morpheme *-in* used with the superlative is used in (23b) with the same effect: pre-nominal ordinal and no Ezafe marker. This provides further support for breaking down the superlative marker into the comparative marker *-tar* and *-in*.

Before we look at Ezafe and Ezafe-like elements in some other Iranian languages, it is worth taking stock of some of the properties of the Persian Ezafe we have seen so far. As highlighted in this paper, Ezafe appears with post-nominal modifiers and never with pre-nominal ones. Moreover, Ezafe appears with adjectives, possessors, as well as PP modifiers, reduced relative clauses and (arguably) with full relative clauses. Meanwhile, Ezafe is not present with superlative adjectives, which are also prenominal. Ezafe never appears on a bare noun, or on a predicative adjective (the latter not shown above). Finally, Ezafe is iterative, in the sense that it appears on the noun when it is modified, and it gets repeated on any additional modifiers except the final one. With this summary, we can now turn to several other Iranian languages.

5. Other Iranian Languages^{33,34}

In this section, we will have a brief look at some Iranian languages other than Persian with respect to the status of Ezafe and Ezafe-like elements. When considering Iranian languages, we come across three types of languages. In the first type of languages, modifiers are prenominal and there is no Ezafe marker. Languages of the second type behave like Persian: they have post-nominal modifiers and the presence of Ezafe. There is yet a third type of language which exhibits a phenomenon termed by some scholar as Reverse Ezafe. This is when there is a prenominal modifier with a Ezafe-like vowel appearing between the modifier and the noun. Let us consider some examples from the three types of languages.

5.1. No Ezafe Languages

This type of language is exemplified here by Pashto, Shughni and Wakhi.

Pashto (East Iranian): Prenominal adjectives with no agreement

- (24) a. *spin motar* b. *agha loy kitaab-un-a*
 white car those big books

(adapted from Robson and Tegey 2012)

Shughni (East Iranian - Pamir): Prenominal adjective with gender agreement

- (25) a. *rosht mun* b. *ter tsimud*
 red(f) apple black basket

(adapted from Edelman and Dodykhudoeva 2012)

³³ I am grateful to Taeho Lee and Sarah Quevedo for their help in collecting the relevant data in this section. Their collaboration with me over the summer of 2015 was funded by two undergraduate University of Toronto Excellence Awards (UTEA).

³⁴ Transcriptions in this section are somewhat loose. For more accurate representations, refer to the original sources.

- (34) az sər-a bar-i vind-əm-a
1Sg. red-a door-Obl. saw-1Sg.-Tr.

‘I saw the red door.’ (cited in Franco, Manzini, Savoia 2015, originally from Paul 2011)

I have not looked closely at Mazandarani or Masali yet, but we carefully reviewed a Grammar of Gilaki (Rastorgueva et al. 2012), sifting through about 150 pages of text. The Gilaki data in (35)-(38) below are all adapted from Rastorgueva et al. 2012.³⁶ We found around 200 noun phrases with nominal or adjectival modification. Over half of them involved the Ezafe construction, N-Ez Adj or N-Ez N, as shown in (35).

- (35) a. utaaq-Ə xaali
room-Ez empty
‘empty room’
- b. sƏrguzƏsht-Ə ita xaanƏvaar-Ə fƏqir
story-Ez one family-Ez poor
‘the story of one poor family’
- c. mu-yƏ siya-yƏ girinji
hair-Ez black-Ez curly
‘curly black hair’
- d. vƏsƏt-Ə taabestaan
middle-Ez summer
‘middle of the summer’

There were about 100 potential cases of REZ. Of these, all except one involved postpositions: N-Ə P; in about half of those the postposition was *miyan* “inside” and the rest other postpositions, namely *durun* “inside”, *ru* “on”, *amara* “with”, *vasi* “for”, *bija* “side, beside”.

- (36) a. aab-Ə miyan c. zahak-Ə amara
water-Ə inside children-Ə with
‘in(side) the water’ ‘with the children’
- b. dƏr-Ə bija d. ab-Ə ja
door-Ə side water-Ə from
‘by the door’ ‘from the water’

This raises the possibility that these are instances of genitive marking, which is how all of them are glossed in the grammar. Adpositions assign genitive case to their nominal complements in other languages, e.g. Arabic. Also, with the nominal status of some prepositions in Iranian languages, this possibility is even more natural. Note that, in Gilaki, unlike Persian, the possessor appears before the possessum and is marked with the same genitive marker. This is true of possessive pronouns as well, as shown in (37).

- (37) a. aahu-yƏ chum-an b. mi xaxur-za
deer-gen eye-plural my sister-child
“deer’s eyes” “my niece”

There was only one example of a prenominal adjective with such marking in the grammar. The authors of the grammar do state that –Ə is productively used to form nouns, adjectives and adverbs. This raises another possibility for the analysis of this vowel.

- (38) pisxaal-Ə rish
small-Ə beard
‘small beard’

³⁶ I am grateful to Taeho Lee and Sarah Quevedo for their help with analyzing the Gilaki data.

Needless to say, this is just a preliminary look at Gilaki based on the data found in one grammar. A closer examination of Gilaki as well as Mazandarani and Masali is needed to gain a deeper understanding of the nature of “Reverse Ezafe” and its relation to Ezafe.

5. Conclusions

It was shown that there is a near-perfect correlation between the order of the noun and other nominal elements and the presence of the Ezafe marker in Persian, with the noun clearly demarcating the distribution of the Ezafe marker: no Ezafe on elements surfacing before the noun and an Ezafe for every element following it. It was argued that these facts are best captured in a system which takes the merge position of the noun in the DP to be final and the surface order derived via roll-up phrasal movement.

We briefly looked at the status of Ezafe and Ezafe-like elements in some other Iranian languages and observed a good range of variation. In order to gain a better understanding of the nature and typology of linker elements used in these and other languages, the relevant data need to be examined much more closely. Crucially, we cannot start with the assumption that these linkers are all the same element in all these languages, and pick properties from different languages to draw unified conclusions about them. Each language should be investigated with respect to all the properties discussed for Persian in this talk. It is conceivable that these elements may have developed different functions in different languages even if they have the same historical source.

References

- Aoun, J., Benmamoun, E., & Sportiche, D. (1994). Agreement, word order, and conjunction in some varieties of Arabic. *Linguistic Inquiry*, 25, 195-220.
- Bashir, E. (2012). Wakhi. In G. Windfuhr (Ed.), *The Iranian languages* (pp. 825-862). New York: Routledge.
- Bassols-Codina, S. (1992). *La phrase relative en kurde central* (unpublished doctoral dissertation). Université de Paris 3, Paris.
- Belleti, A. (2004). *Structures and beyond: The cartography of syntactic structures*. Vol. 3. Oxford: Oxford University Press.
- Cheng, L. (1997). *On the typology of wh-questions*. New York: Garland.
- Chomsky, N. (1995). *The minimalist program*. Cambridge, MA: MIT Press.
- Cinque, G. (1994). On the evidence for partial N movement in the Romance DP. In G. Cinque, J. Koster, J.-Y. Pollock, L. Rizzi, & R. Zanuttini (Eds.), *Paths towards universal grammar: studies in honor of Richard S. Kayne* (pp. 85-110). Washington, DC: Georgetown University Press.
- Cinque, G. (1996). The ‘antisymmetric’ programme: Theoretical and typological implications. *Journal of Linguistics*, 32, 447-464.
- Cinque, G. (1999). *Adverbs and functional heads: A cross-linguistic perspective*. Oxford: Oxford University Press.
- Cinque, G. (2002). *The functional structure of DP and IP: The cartography of syntactic structures*. Vol. 1. Oxford: Oxford University Press.
- Cinque, G. (2005). Deriving Greenberg’s Universal 20 and its exceptions. *Linguistic Inquiry* 36: 315-332.
- Cinque, G. (2010). *The syntax of adjectives: A comparative study*. Linguistic Inquiry Monograph 57. Cambridge: MIT Press.
- den Dikken, M. (2006). *Relators and linkers: The syntax of predication, predicate inversion, and copulas*. Linguistic Inquiry Monograph 47. Cambridge: MIT Press.
- Edelman, D. I., & Dodykhudoeva, L. R. (2012). Shughni. In G. Windfuhr (Ed.), *The Iranian Languages* (pp. 787-824). New York: Routledge.
- Ernst, T. (2001). *The syntax of adjuncts*. Cambridge: Cambridge University Press.
- Franco, L., Manzini, M. R., & Savoia, L. M. (2015). Linkers and agreement. *The Linguistic Review*, 32(2), 277-332.

- Ghaniabadi, S. (2010). The Empty Noun Construction in Persian (unpublished doctoral dissertation). University of Manitoba, Winnipeg.
- Ghomeshi, J. (1996). Projection and inflection: A study of Persian phrase structure (unpublished doctoral dissertation). University of Toronto, Toronto.
- Ghomeshi, J. (1997). Non-projecting nouns and the Ezafe Construction in Persian. *Natural Language and Linguistic Theory* 15, 729-788.
- Halle, M., & Marantz, A. (1993). Distributed Morphology and the Pieces of Inflection. In K. Hale, & S. J. Keyser (Eds.), *The view from Building 20* (pp. 111-176). Cambridge: MIT Press.
- Holmberg, A., & Odden, D. (2005). The Izafe and NP structure in Hawrami. *Durham Working Papers in Linguistics* 10, 77-94.
- Irimia, M., Longobardi, G., Michelioudakis, D., & Radkevich, N. (2014). Case, PRO-features, and nominal linkers: Evidence from Romanian genitives. Paper presented at Going Romance, Lisbon.
- Kahnemuyipour, A. (2000). Persian Ezafe Construction revisited: Evidence for Modifier Phrase. In J. T. Jensen, & G. van Herk (Eds.), *Cahiers Linguistique d'Ottawa* (pp. 173-185), proceedings of the 2000 annual conference of the Canadian Linguistic Association.
- Kahnemuyipour, A. (2014). Revisiting the Persian Ezafe Construction: A roll-up movement analysis. *Lingua*, 150, 1-24.
- Karimi, S., & Brame, M. (2013). A generalization concerning the EZAFE Construction in Persian. *Linguistic Analysis*, 38, 111-143.
- Karimi, Y. (2007). Kurdish Ezafe Construction: Implications for DP structure. *Lingua* 117, 2159-2177.
- Kayne, R. S. (1994). *The antisymmetry of syntax*. Cambridge: MIT Press.
- Larson, R. (2009). Chinese as a Reverse Ezafe language. *Yuyanxue Luncong (Journal of Linguistics)*, 39, 30-85.
- Larson, R., & Yamakido, H. (2006). Zazaki "Double Ezafe" as double case-marking. Paper presented at the Linguistics Society of America annual meeting, Albuquerque, New Mexico.
- Larson, R., & Yamakido, H. (2008). Ezafe and the deep position of nominal modifiers. In L. McNally, & C. Kennedy (Eds.), *Adjectives and adverbs: Syntax, semantics and discourse*, (pp. 43-70) Oxford: Oxford University Press.
- Lazard, G. (1992). *A grammar of contemporary Persian*. Costa Mesa, California: Mazda Publishers.
- Marantz, A. (1997). No escape from syntax: Don't try morphological analysis in the privacy of your own lexicon. In A. Dimitriadis, L. Siegel, C. Surek-Clark, & A. Williams (Eds.), *U Penn Working Papers in Linguistics* 4.2 (pp. 201-225). Philadelphia: Penn Linguistics Club.
- Matushansky, O. (2008). On the attributive nature of superlatives. *Syntax* 11, 26-90.
- McGinnis, M. (2013). Irreversible binding. In S. Kan, C. Moore-Cantwell, & R. Staubs (Eds.), *Proceedings of NELS* 40 (pp. 81-92). UMass Amherst: GSLA.
- Moyne, J., & Carden, G. (1974). Subject reduplication in Persian. *Linguistic Inquiry* 5(2), 205-249.
- Paul, D. (2011). *A comparative dialectal description of Iranian Taleshi* (unpublished doctoral dissertation). University of Manchester, Manchester.
- Pearce, E. (2002). DP structure and DP movement in Maori. Paper presented at COOL5, University of Canberra, Australia.
- Rastorgueva, V.S., Kerimova, A.A., Mamedzade, A.K., Pireiko, L.A., & Edel'man, D.I. (2012). *The Gilaki Language* (English translation and expanded content by R. M. Lockwood). Acta Universitatis Upsaliensis, Studia Iranica Upsaliensia 19. Uppsala University.
- Ritter, E. (1991). Two functional categories in noun phrases: Evidence from Modern Hebrew. *Syntax and Semantics*, 25, 37-62.
- Ritter, E. (1992). Cross-linguistic evidence for Number Phrase. *Canadian Journal of Linguistics*, 37(2), 197-218.
- Rizzi, L. (2003). The structure of CP and IP: The cartography of syntactic structures. Vol. 2. Oxford: Oxford University Press.
- Robson, B., & Tegey, H. (2012). Pashto. In G. Windfuhr (Ed.), *The Iranian Languages* (pp. 721-771). New York: Routledge.
- Samiian, V. (1983.) Origins of phrasal categories in Persian, an X-bar analysis (unpublished doctoral dissertation). UCLA. Los Angeles.
- Samiian, V. (1994). The Ezafe Construction: Some implications for the theory of X-bar syntax. In M. Marashi

- (Ed.), *Persian Studies in North America* (pp. 17-41). Bethesda, MD: Iranbooks.
- Samvelian, P. (2007). A (phrasal) affix analysis of the Persian Ezafe. *Journal of Linguistics*, 43, 605-645.
- Samvelian, P. (2008). The Ezafe as a head-marking inflectional affix: Evidence from Persian and Kurmanji Kurdish. In S. Karimi, V. Samiiian, & D. Stilo (Eds.), *Aspects of Iranian Linguistics* (pp. 339-361). Newcastle upon Tyne: Cambridge Scholars Publishing.
- Shlonsky, U. (2004). The form of Semitic noun phrases. *Lingua*, 114.12, 1465-1526.
- Skjærvø, P. O. (2009). Middle West Iranian. In G. Windfuhr (Ed.), *The Iranian Languages* (pp. 196-278). New York: Routledge.
- Toosarvandi, M., & van Urk., C., (2014). The syntax of nominal concord: What Ezafe in Zazaki shows us. *Proceedings of NELS*, 43(2), 209-220.

a.kahnemuyipour@utoronto.ca

The acquisition of “other” “different” “less common” “distant” languages: A critical need

Silvina Montrul
University of Illinois

Abstract

Linguists compare different languages to formalize the systems that govern language, and the goal of language acquisition research is to understand the processes by which the systems that govern language grow in the minds of language learners. However, much of linguistics and language acquisition theorizing has been primarily built on the empirical foundation of English, and a few other European languages. But the complexity and universality of language as well as the processes underlying its acquisition in monolingual and multilingual situations necessitates a wider and more diverse empirical base. In this talk, I will discuss how results from the acquisition of languages such as Turkish, Hindi, Romanian, Spanish and Farsi, among others, contribute to theory building and to inform pedagogical practices of second, third and heritage language acquisition. Ultimately, I argue that learning and doing research on such languages preserve and enhance linguistic diversity and, ultimately, foster greater cultural and political understanding.

Refining Turkish Stress as a Multifaceted Phenomenon

Bariş Kabak
University of Würzburg, Germany

Abstract

Much has been said about stress assignment in Turkish, albeit with dubious characterizations of its nature and dynamics in the realm of prosodic typology and psycholinguistics. On the basis of crosslinguistic and experimental evidence, I will unpack the notion of “accent” at the level of the word and argue that word stress in Turkish is inert but pervasive, exhibiting both a demarcative and a lexical character, tacked to a relatively poor pitch-based intonational inventory. Due to morphosyntactic complexity of words in Turkish, I will argue that accentual phenomena are ultimately intertwined with wordhood, whereby the induction of canonical right-edge stress in language acquisition must be word-based and paradigmatic. I will further show that word formation has also been the impetus behind the morphophonologization of stress diachronically, ultimately surmounting words with accentual prominence. Potential venues for the investigation of word prosody such as text-to-tune alignment in lyrics will be highlighted.

How did the Turkic languages come to differ as they do?

Marcel Erdal
Goethe University, Germany

Abstract

After more than two centuries of increasingly intensive research on these languages, the field still offers a great many of problems to be solved and sources for building or testing theories. These languages that form a small family show manifold outer contacts and far-reaching influence. The issues to be discussed include typology (CSCP or “left-branching” syntax, mostly “right-branching” agglutination; vowel harmony, phonotactics); disputed classifications of the living languages and dialects; periodization; “Mongolic” and “Para-Mongolic”; Mongol influence on other languages; influence of other languages on Mongol; two millennia of the Altaic Sprachbund; questions to be answered, fields to be explored.

1. Introduction

Close to 40 Turkic languages are spoken between Heilongjiang in the East, Poland and the Balkan peninsula in the West, the Laptev Sea in the North and the Persian Gulf in the South. This huge spread explains the lexical differences between them, with Russian loans dominating in the North, Chinese ones in the East, Arabic and Persian ones in the Muslim Turkic languages and Mediterranean loans in the South West; Mongolic lexical inroads are found everywhere, but more in the East than in the west – a legacy of Genghis Khan’s empire. We will not deal with this fascinating Eurasian cultural interaction in detail, as all such lexical wealth is now accessible on the internet: Luckily, practically all Turkic languages are exclusively suffixing,¹ so that any inflected word can be easily looked up.

The syntax of Turkic languages is quite similar, with left-branching subordination retained as the canonical structure in all of them; inside the noun phrase, inverse order is permitted only in one or two non-Muslim languages in the far west, which show strong Slavic influence. Infinite verb forms serve clause subordination in all Turkic languages; they differ in the way the subject of subordinate clauses is expressed. However, since the 9th century, practically all documented Turkic languages have also had, due to Iranian or Slavic influence, right-branching relative, complement and adjunct clauses with finite verbs, introduced either by conjunctions or inflected pronouns. Only very few Turkic idioms, e.g. Sayan Turkic varieties spoken in Mongolia, fully adhere to the canonical Turkic type of clause subordination. The often-mentioned S – O – V order of sentence units is not absolute in any Turkic language past or present; it is modified in all Turkic languages for pragmatic purposes, which are, after all, the main reason for utterances in the first place. Turkic languages differ in the degree to which auxiliaries serve the expression of actionality, Turkish appearing to be the Turkic language in which they are least used. Otherwise, knowing any one Turkic language, one will not be surprised by the syntactic structures of any other Turkic language one comes across.

The differences between the Turkic languages which will cause the greatest problems to learning are in the phonic and the morphological domains. To master these differences, the present paper proposes a historical approach: If we can describe Proto-Turkic, spoken perhaps two and a half millennia ago, and sketch all the phonological and morphological changes which took place in all languages coming from it, we will get the sound and word structures of all the Turkic languages. Before coming to the question of how it all started and before looking at how the different features of the earliest stages of Turkic evolved to give the present-

¹ The only exception is Uyghur, where stems can be changed through leftward harmony. We will come to that below. The only non-suffixing process is adjective doubling as in Turkish *sap-sarı* ‘bright yellow’ < *sarı* ‘yellow’, which is easily learned. Pronouns can be quite irregular and their forms have to be specially acquired.

day languages, we will first mention all the Turkic languages and see how they relate to each other within the genetic groups. Languages change through divergence and convergence. Divergence is when languages split into dialects which then come to differ so much that they become different languages. This happens when societies split and the different branches go to live in different places and are no longer one society; usually they in those places also have contact with speakers of different other languages. One example for this would be speakers of Chaghatay Turkic, which served as the standard language of the Timurid realm in the 15th century; it turned into Modern Uyghur in the east through intensive contact with Mongolic and Chinese, but turned into Modern Uzbek in the west through intensive contact with Persian and Kipchak Turkic and more recently Russian. We mentioned the contact with Kipchak Turkic: This is a case of convergence: Modern Uzbek resulted from the convergence of Chaghatay Turkic with the steppe variety of Kipchak Turkic. Another interaction of linguistic divergence and convergence resulted in the emergence of the Turkmen language: The Turkmen were the members of the heathen Oghuz tribes living north-east of the Caspian Sea in the 10th century, when Selchük, a warlord, became Muslim with his men, moved south and began fighting against them. The Selchük empire moved to modern Iran, Iraq, Syria and Turkey. This rift caused a divergence between the non-Selchük and Selchük Oghuz varieties. The language of the non-Selchük Turkmen, who stayed in Central Asia, subsequently converged with both Kipchak Turkic and Chaghatay Turkic. To understand how the Turkic languages changed, we need to describe the divergences of their genetic clusters in the first place, and supplement this view with the geographic distribution, which caused their convergences. We will first list members of each genetic group, then mention some prominent features which distinguish it from the other groups. Generally (though not necessarily universally) shared features will not be mentioned: e.g. The form with $-(X)p$,² used as converb and for coordinative linking, e.g., need not be mentioned when dealing with Oghuz, as all Turkic languages except Chuvash and Yakut have it; nor need $+dA$ be mentioned, as it is used for the locative everywhere except in Yakut (where it instead forms partitives).

2. A genetic survey of the Turkic languages, and some of their distinctive features³

Since we have just mentioned the Oghuz, we might as well start this survey with Turkey, where Oghuz nomads arrived 1000 years ago, even before the Selchüks conquered Baghdad in 1055 and before they beat the Greeks in Eastern Anatolia in 1071. Beside the Turkish population of Turkey, **Western Oghuz** comprises Turkish dialects in Bulgaria, ex-Yugoslavia, North-Eastern Greece, Cyprus and on the island of Rhodes. The Gagauz, a Christian people mostly living in Moldova, a country north of Rumania, as well as in Bulgaria, have a Balkan-Turkish standard language which appears to have split from Western Oghuz at an early stage. Azeri, the only written language of **Central Oghuz**, is spoken in Azerbaijan and, in a strongly Iranized variety, in Iran (26% of the population of Iran is Turkic speaking). The linguistic transition in Anatolia towards the east is gradual, with eastern Anatolia varieties closer to Azeri than to Standard Turkish. Meskhetian Turkish is spoken in villages in Georgia and in a diaspora in Kazakhstan. The Turkmen in northern Iraq do not speak Turkmen; they speak a variety somewhere between Azeri and Turkish, already used by the 16th century poet Fuzuli, born in Baghdad. There are more than a million Qashghay in South-West Iran, and others speaking the Aynallu, Sonqori, Ghalûgha and other varieties of Middle Oghuz. **South Eastern Oghuz** is Khorasan Turkic, spoken in the east and north-east of Iran; the Oghuz varieties spoken in western Uzbekistan are close to it. We already mentioned the great rift dividing the Selchük Oghuz from the original Oghuz tribes in present-day western Kazakhstan, who became the Muslim Turkmen only in the 14th century. Under Kipchak pressure, they moved south to conquer all of present-day Turkmenistan, and reached north-western Afghanistan and north-eastern

² Brackets mark segments which are retained under certain phonic circumstances, dropped in others. Capital letters represent sounds which alternate because of vowel harmony or assimilation: *U* is *u/ü*, e.g., *D* is *t/d*, *X* *i/i/ü/u*. + marks nominal, - verbal juncture. The colon (:) shows vowel length. The forms given are those of the spoken language, since spelling often masks reality.

³ This section is based on the work of Claus Schönig; I dedicate the paper to him.

Iran as well. This migration separated the Khorasan Turkic speaking groups in Uzbekistan and in Eastern Iran. The Turkmen are the real, pure Oghuz, the only ones who kept their original tribal structure and, till the 20th century, their nomadic life style. However, staying in north-western Central Asia, **North-Eastern Oghuz** grammar was influenced on the one hand by Chaghatay and on the other hand by Eastern Kipchak, to which we come next. Selchük Oghuz, on the other hand, became part of the Near East.

Features distinguishing Oghuz from the other Turkic languages are the retention of the vowel length opposition (in Central and Western Oghuz only in reflexes), the change of original /ñ/ (which did not exist as such in any contact language) to /yn/ or /yVn/ and the change of /d/ after vowels to /y/. Only Western Oghuz retains onset /b/ in words such as *bän* ‘I’ or *bin-* ‘to mount, to ride’, where the second consonant is a nasal; such words start with /m/ in the other languages. In nominal morphology there is the genitive suffix *+(n)Iŋ*, which drops its onset /n/ after consonants, an archaic feature even in Old Uyghur, otherwise found only in Chuvash. The accusative suffix is *+(y)I* only in Western Oghuz; the other Oghuz languages have *+(n)I*, most other Turkic languages *+nI*. The *+rA+* extension of pronominal stems in *orada* etc. is purely Oghuz (although Proto-Turkic did have a local case suffix of this shape). An evidential verbal suffix *-mIš⁴* survives outside Western and Central Oghuz in this function in Yakut, Salır and Khalaj (though some other languages did retain a dubitative particle *imiš*), nor is there a participle suffix *-mIš* in other languages. The use of *-(y)An* for subject participles and *dIk+* for participles with subject suffixes is only found in Khalaj, in Western and in Central Oghuz, not even in Turkmen: Turkmen has developed an opposition between an *-An* participle for past events and a *-yAn* participle for ongoing events. In this form, in the accusative suffix mentioned above, in the 1st person clitic pronouns, in the *-(y)Xp* converb and elsewhere, Central and Western Oghuz make wide use of a hiatus-bridging consonant /y/. The suffix referring to 1st person plural verbal subjects, which originally was *-(X)mXz* like the possessive suffix, was changed to *-k*. Oghuz on the other hand preserved *+(I)z* in 2nd person plural verb forms, in most Turkic languages replaced by *+lAr*. The negative converb suffix *-mAyIp* is a western Oghuz innovation; others have *-mAy*, *-mAyIn*, *-bA:n* etc. (from earlier *-mAtIn* and *-mAdIn*). The Oghuz nominal negation comes from *tägül*. Only Turkish and Yakut have preserved the reflexive pronoun *käntü*, only Turkish has given the verb *yap-*, originally ‘to cover’, the meaning ‘to do’. Turkmen differs from the other Oghuz languages in another important point: No /y/ is placed between vowels at the end of stems and vowels starting suffixes, as happens e.g. in the dative form; instead, the two vowels fuse into a long vowel. Oghuz uses a suffix coming from the vowel converb plus *yorr-* ‘to walk’ for expressing continuous content. Finally, the *-AmA-* form expressing inability is Western Oghuz; in the other Oghuz languages we find *-bilme-* in this function, while other Turkic languages use *al-* for expressing ability and inability. In many of these matters, Khorasan Turkic is intermediate between Selchük Turkic and Turkmen.

Many scholars think that the language of the **Salır**, who live in the Qinghai and Gansu provinces of China, also came from the Oghuz group; there is also a Turkmen dialect Salır, coming from Salgur, the name of an 11th century Oghuz tribe. They base their view on the fact that Salır has reflexes of long vowels, on its finite *-mIš* form, on Salır *eyle* which echoes Oghuz *öyle* ‘thus’, not found in any other branch of Turkic (other Turkic languages have forms coming from *antag*), and a number of lexemes typical of Oghuz. The Salır genitive suffix is *+niyi* (< **-nIŋ+ki* with generalized function, as in the Lena Turkic personal pronouns), not dropping the onset /n/. Like Oghuz (and Chuvash), its dative suffix is *+A* after consonants; but *+GA*, which it has after vowels, is like all the other Turkic languages. Salır has, like the Siberian languages, Chuvash and Yellow Uyghur, retained *-dImIz* as 2nd person past tense form, and not changed it to *-dIk*. Like the Siberian languages and Yellow Uyghur, it lacks **tägül* ‘is not’ and the privative suffix *+sIz*. In the 2nd person plural verb forms Salır adds *-lAr* after *-Xz*. Summing this up we note that, if Salar was an Oghuz language at some stage, it lacks some of the typical features of that group.

⁴ š is the voiceless palatal sibilant, ž the voiced palatal sibilant, č the voiceless palatal affricate, ĵ the voiced palatal affricate.

Having mentioned Iran, one might here add **Khalaj**, a language spoken in some villages in central western Iran which constitutes an independent and archaic branch of the Turkic languages. Khalaj has been connected with the Early Turkic Argu dialect referred to in the Dîvân Lughâti 't-Turk, e.g. in preserving the element *day* for 'is not' and in turning **ñ* to *n*. It is more archaic than Old Uyghur in using *+(U)ŋ* as genitive suffix, in preserving the length distinction in vowels and the sound /h/ in word onset, e.g. in *higla-* 'to weep' or *hagač* 'tree'; this feature survives in Old Turkic only in reflexes. Old Turkic /X/, the fourfold vowel harmony, corresponds to Khalaj /U/, which appears also in early Turkic loans in Mongolic; the abstract suffix, e.g., is *+Iuk*. It is the only language to preserve the Old Turkic dative suffix as *+kA* and the ordinal suffix as *+(I)nč*. *kem* 'who' is also archaic, and Khalaj has preserved Proto-Turkic and Old Turkic /d/ after vowels and /g/ in all positions. The relative participle in *-dUk* with possessive suffixes referring to the subject, the present tense form coming from *-A yor-*, the *-AĵAk* future, the converb form ending in *-ArAk*, interrogative-indefinite pronouns coming from Old Turkic *kañu* are among the features exclusively shared by Oghuz and Khalaj.

We now turn to the **Kıpçak** languages, which can be divided into several groups. We have the **Volga-Ural** group, consisting of Tatar, spoken in Tatarstan (Volga-Kama region) and in regions around it, in the ex-Soviet republics and in China, and of Bashkir. Misher Tatar is an important dialect spoken to the west of standard Tatar, while Tatar dialects in Siberia, the Tümen, Tobol, Baraba, Tom and a number of other dialects, form a linguistic bridge to Siberian Turkic. The **Caucasus Kıpçak** languages Karachay-Balkar and Kumik have been influenced by Caucasian languages. **Karaim**, strongly influenced by Slavic, was until recently spoken in the Ukraine and is still spoken in Poland and Lithuania; Misher Tatar shares some features exclusively with it. **Crimean Kıpçak** consists of Crimean Tatar and of non-Muslim dialects close to it. It was – and is again – spoken in the Crimean peninsula, in Kazakhstan, where Stalin deported its speakers, and in Rumania. Of Crimean Karaim and Crimean Krimchak, the Jewish varieties of Crimean Tatar, the first appears to have died out, while there still are some speakers of the latter. There are two Turkic varieties called Urum because the speakers are ethnic Greeks who have also emigrated to Greece and Cyprus; there is a Kıpçak variety spoken north-east of the Crimea and an Oghuz variety spoken in Georgia. We spoke of the Turkmen as the 'original' Oghuz who kept their tribal structure and nomadic life style. Among the Kıpçak, the Kazakhs are those who kept their tribal structure and nomadic life style, but the reason for this may also be the ethnical admixture of Mongol tribes among them. Kazakh is quite homogenous in a huge area stretching from the Volga all the way into China; in the Kazakhstan - China - Mongolia - Siberia border area, it is spoken in all four countries. Karakalpak, spoken in the northwestern part of Uzbekistan, and Kazakh are members of the **Nomadic Kıpçak** group. Noghay, spoken in various places north of the Caucasus, north of the Crimea and in Rumania, has some Caucasus and Western Kıpçak features, in a number of ways being intermediate between these languages. **Kırgız**, which also shares some features with Kazakh, is considered to be Kıpçak as well, though it differs in some important ways, bringing it close to Altay Turkic. Altay Turkic is thus intermediate between Kıpçak and Siberian Turkic.

All Kıpçak languages lost /g/ after vowels, but this in some cases led to diphthongs with /w/ and in other cases the loss resulted in the lengthening of the preceding vowel. They have no reflex at all of Proto-Turkic vowel length, though the Volga-Ural languages have a contrast of normal and reduced vowels as an areal feature shared with Finno-Ugric languages and Chuvash, coupled with a vowel shift. Kıpçak languages voice consonants between vowels whereas Central and Western Oghuz does this only after originally long vowels. Onset /y/ often appears as /j/, and onset /b/ becomes /m/ if the following consonant is a nasal. The accusative suffix is *+nI*, the dative suffix *+GA*. The onset of the genitive suffix **+nIŋ* is never dropped and Kıpçak languages have replaced evidential *-mIš* with analytical phrases. The *-mIš* participle has been replaced by *-gAn*; when it qualifies heads which are not its subjects; subjects of subordinate clauses are referred to by independent pronouns and not by possessive suffixes. There is a modal form in *-gAy*, expressing the speaker's attitude towards possible future events. The Kıpçak languages are otherwise similar to the Oghuz languages in having Proto-Turkic /d/ represented as /y/ after vowels, *kim* for 'who', *-k* as 1st person plural verbal suffix and *öz* as reflexive pronoun. The 1st person possessive suffix can be *+(I)bIz* or *+(I)mIz*, the 2nd person *+(I)gIz*

or $+(I)\eta Iz$. The European Kıpçak languages have forms of *tägül* for ‘is not’, but the Central Asian Kıpçak languages use forms coming from *ärmäz* (Kırghız has both). The European Kıpçak languages have $-Iz$ in the 2nd person plural verb forms, but we find both $-Iz$ and $-IAr$ in the Central Asian Kıpçak languages. The present tense is formed by $-(y)A$ followed by personal pronouns (sometimes shortened to $-m$, $-s$, $-t$). Kıpçak languages are unlike Oghuz in using suffixes coming from *tur-* ‘to stand’ for expressing continuous actionality; only Kazakh also uses *yür-* ‘to walk’, *otır-* ‘to sit’ and *yat-* ‘to lie’ for this content, as the eastern and Siberian Turkic languages do. *al-* ‘to take’ has come to express possibility, partly as suffix $-(y)al-$ or even $-(y)Al-$.

We now come to **Uzbek**, whose standard variety and main dialects have lost the vowels /ö ü ı/ and vowel harmony in general, through a millennium of Persian influence: Much of the present-day population of Bukhara and Samarkand actually speaks Tajik, a dialect of Persian with Turkic adstrate. Other results of Persian influence are the labialization of much of the /a/ phoneme (borrowed /â/ and accented Turkic /a/, spelled o) and the use of numerous Persian derivational elements. Uzbek *with* vowel harmony and the vowels /ö ü ı/ is spoken in southern Kazakhstan. Millions of Uzbek speakers live in northern Afghanistan, there is a sizable Uzbek minority in southern Kırghızstan and some 2 million Uzbek in Russia and elsewhere. On the other hand, we already mentioned that Oghuz Turkic is spoken in south western Uzbekistan, and Karakalpak, akin to Kazakh, in the north west; Kazakh is spoken elsewhere in Uzbekistan close to the Kazakh border. Uzbek is essentially a continuation of Middle Turkic Chaghatay with Kıpçak influence: Thus, the Old Turkic suffix $+IXg$ became $+li$, the postposition *täg* became *däy*, not $+Ilk$ and *däk*, as in Modern Uyghur. Similarities to Modern Uyghur: Proto-Turkic /d/ becomes /y/ after vowels. Unlike neighboring Kazakh and Kırghız, consonant assimilations are limited to voicing, the onset consonants of the case suffixes $+nI$ (accusative), $+nI\eta$ (genitive) and $+GA$ (dative) are not dropped after consonants and the 3rd person possessive suffix gets no /n/ when followed by case suffixes. *siz* is the normal address also for any single addressees. ‘Which’ is *qa(y)+* (as in Kıpçak), ‘is not’ is *emäs*. Relative clauses are formed around the $-GAN$ participle (past and general present), expanded with *yat-* ‘to lie’ to give *yätkän* (immediate present) and with *tur-* ‘to stand’ to give *digän* (continuous and future); if the head is not the subject of the relativized verb, the subject is referred to either by a possessive suffix on the head (not on the verb form) or with an independent pronoun. Very many verbs are used as auxiliaries for expressing actionality, either with the $-A/-y$ or with the $-(i)p$ converb. Ability is expressed by *al-* ‘to take’, which becomes a suffix with vowel harmony in Uyghur. The 3rd person clitic pronoun is *dı* in Uzbek, *du* in Uyghur ($< tur-ur$).

Modern Uyghur, mostly spoken in Xinjiang in western China, also lost the phonemic opposition between /i/ and /ı/ but does not labialize /a/s and has retained /ü/ and /ö/; the harmony class in /U/ is not fused with /I/ or /X/, e.g. in the verbal noun suffix $-gu \sim -gü$. A few Proto-Turkic long vowels are preserved. /r/ is dropped after vowels but can appear between vowel coda and vowel onset also where not justified etymologically. I just mentioned some of the Uyghur features when discussing Uzbek; a further development is the fusion of $-(i)p yat-$ to give $-(i)wat-$ as marker of continuous aspect. Uyghur grammar is a direct continuation of Chaghatay, but with a Mongolic substrate. Unlike Uzbek, all coda /g/s get devoiced to /k/ and there is raising and fronting of vowels through backward assimilation. Backward fronting makes /a/ to /e/ and not /ä/, giving an eighth vowel phoneme. The ablative suffix is $+din$ as in Old Uyghur, not $+dAn$ as in Uzbek. The 2nd person plurals *sän* and *sänlär* are so familiar as to be rare; *siz* and *silär* are the normal addresses to everybody. *alli+* (from Mongolic) is prefixed to interrogative pronouns to form indefinites. $(i)miš$ can be added to any sentence to express evidentiality but there are analytical means as well. Finite verbs are pluralized with $-IAr$. The modal use of the $-gAy$ form, expressing the speaker’s attitude towards possible future events, differs from its indicative use in Old Uyghur. Relative and complement clauses use the perfect participle and verbal noun in $-gAn$, the progressive one in $-(i)watqan$ or the general and prospective one in $-(i)diyan$ with subjects referred to by independent pronouns in relativization (but possessive suffixes if there is no head) and possessive suffixes in complementation. The purpose meaning of $-gAll$ forms and the participle in $-gUči$ are, beside Uyghur, preserved only in Khalaj.

In Siberia there are four Turkic branches: **Altay Turkic** in central South Siberia, **Yenisey Turkic** further East, **Sayan Turkic** north of Mongolia and **Lena Turkic** in northern Siberia. These are different genetic

branches, but there has been some convergence among the south Siberian ones. **Altay Turkic** consists of the language of the Altay Kizhi, of Telengit and of Teleut. **Yenisey Turkic** consists of Khakas (which has several distinct dialects), of Shor, of Chulim Turkic (one dialect changing **d* to *y*, the other to *z*) and of the dialects of the Kumandi, Chalkandu and Tuba living in the north of the Altay Republic and now under strong Altay Turkic influence. **Yellow Uyghur**, spoken in the Chinese province of Gansu, is close to Yenisey Turkic although it also has some independent features; so was **Fu-Yü 'Kırgız'**, spoken till a few years ago in Heilongjiang, i.e. Manchuria. The standard language of **Sayan Turkic** is Tuvan; other varieties are Tofa, Toju and (in Buryatia, close to the Baykal lake) Soyot. Dukha, spoken in Mongolia, belongs to the same group, as do other Tuvan dialects in Mongolia. The Diva in the northernmost tip of Xinjiang (China) also speak Sayan Turkic. **Lena Turkic** consists of Yakut (endonym Sakha) and Dolgan (which could also be considered a dialect of Yakut). None of the ethnic groups mentioned in this paragraph is Muslim; as a result, Arabic and Persian loans are quite few here, though not completely absent. In Lena Turkic, the harmony of low vowels gives /O/ when following /O/, as in Altay Turkic, Kırgız, the Lop Nor dialect of Modern Uyghur and (to judge by sources in Brâhmî and Tibetan writing) Old Uyghur. Lena Turkic has fully retained the Proto-Turkic vowel length distinction (though the low long vowels here appear as falling diphthongs); Sayan Turkic and Yellow Uyghur retain reflexes of them but the other Siberian languages have lost the distinction. Those reflexes consist of aspiration in Yellow Uyghur and of glottalization in Sayan Turkic, characterizing the vowels which originally were not long. This glottalization has been ascribed the Samoyed influence, as have the change of onset /y/ to /n/ if the next consonant is an /n/, and the sound change *-w > -g*. In all of Siberian Turkic, the 1st person possessive suffix has lost its nasality and become *+(l)bls*; in Lena Turkic, the /s/ further became /t/ (as did coda /s/ and /z/ in general). The Lena Turkic 3rd person possessive suffix is *+(t)A*; we just mentioned that /t/ < /s/ is regular, but the low vowel not found anywhere else. No Siberian language has **tägül* (but note Khakas *çoyıl* < *yok ol*, similar to other fused sequences of *ol*); its Sayan Turkic counterpart comes from *ärmâz* while Yenisey Turkic has added *âmäs* to some pronominal element, giving *nimäs*. As for the /d/ criterion, Proto-Turkic /d/ after vowels stayed unchanged in Sayan Turkic, became /t/ in Lena Turkic and /z/ in Yenisey Turkic including Fu-Yü 'Kırgız' and Yellow Uyghur (but one Chulim Turkic dialect and one Shor dialect have /y/ instead). We see /y/ in Altay Turkic, which also in other ways is close to the Kıpçak branch, e.g. in its loss of /g/ after vowels. Yenisey and Sayan Turkic have retained all such /g/s while Lena Turkic also lost them. The Lena Turkic accusative suffix is *+(n)I* as in Azeri and Turkmen; it has no genitive suffix at all (the stem form is used as genitive, its function made explicit by the possessive suffix on the head) but the South Siberian languages have the full form of this suffix. The Lena Turkic dative suffix also has locative meaning, under Mongolic influence. 'who' is *käm* in Altay and Sayan Turkic, in Fu Yü and in Yellow Uyghur, but in Sayan Turkic has high vowels. The grammaticalization of *öz* to become the reflexive pronoun did not reach Siberia. Lena Turkic (like Western Oghuz) preserved the original reflexive pronoun, *käntü*, as a demonstrative, and borrowed Mongolic *bäyä* 'body; person' as reflexive. Yellow Uyghur also has this, but the normal reflexive pronoun is there another Mongolic borrowing, *ejen* 'master'. In the Yenisey and Sayan Turkic languages, *bod* 'the human stature' was made into reflexive pronoun as a loan translation from Mongolic. The ordinal suffix is common Turkic *+(X)nčI* in Yenisey Turkic, but Lena Turkic has *+(X)s* and Sayan Turkic an expansion, *+(X)škI*. Yenisey Turkic, Fu Yü and Altay Turkic are like common Turkic in having adopted *et-* for 'to do' whereas Sayan and Lena Turkic, beside *kıl-*, also have a variant *kın-* with this meaning. All Siberian Turkic languages as well as Fu Yü and Yellow Uyghur have kept the possessive suffix in the 1st person plural of the past tense. Lena Turkic has kept *-Iz* in 2nd person verb forms but the South-Siberian languages switched to *-IAr* also for the verb; it also kept the finite use of **-mlš* (> *-blt*) but the South-Siberian languages lost it. Lena Turkic, like Khalaj and Chuvash, lacks the otherwise highly active *-(X)p* converb; this form appears to have become important after the ancestors of the speakers of these languages left the Turkic community.

3. Proto Turkic

We were going to consider the reconstruction of the source of all Turkic languages before they came to differ. I assume that a more or less homogenous Turkic language was spoken perhaps in the 5th century B.C.,⁵ perhaps in Western Mongolia and in the South Siberian regions just north of that: It split into Common Turkic and Proto-Bulgar in the 3rd century B.C., when Proto-Bulgar⁶ was exerting much linguistic influence on Proto-Mongolic. After that, we take the Proto-Bulgars to have been defeated by the Xiongnu and to have migrated west, to central-western South Siberia. There they had reasonably datable contacts with the Proto-Samoyeds, a Uralic people, before moving further west.

The oldest surviving Turkic texts are the inscriptions written in the runiform alphabet; the few monuments in the Orkhon and Selenge valleys, datable to the start of the 8th century, have rightfully attracted most of the attention, but there are in fact several hundred rock surfaces and objects with undated runiform inscriptions all over Mongolia and South Siberia, as well as in Kirgizstan and China. We have no direct evidence for the language spoken among the elite of the 1st Türk qaghanate established in the middle of the 6th century; both inscriptions pertaining to it are written in Sogdian, an Iranian language. Its titles and the proper names of its rulers are not interpretable as Turkic.⁷

For some reason, some people wrongly think that the language of the inscriptions of the 2nd qaghanate is Proto-Turkic: In fact, other Turkic tribes like the Basmil, Karluk, Bayırku, Kurikan, Türgäş, Tölis or Tarduş are mentioned in these inscriptions, and they probably spoke different dialects. One can assume that the speech of the tribes which became part of the Türk empire converged to some degree, but not all Turks were part of it. Maḥmūd ibnu 'l-Ḥusayn, known as Kāshgharī, the great 11th century scholar, gives us linguistic evidence for 20 different tribes and for 7 cities, one of them his native Barsghan near the Isık Kōl in present-day Kirgizstan. I will mention just a single feature which distinguished Proto-Turkic dialects, the shape of the dative suffix. In practically all of Old Turkic, i.e. in the runiform inscriptions, in Old Uyghur and in Muslim Khâqânî Turkic, the dative suffix is +*kA*; we find +*gA* only in a few Manichaean manuscripts and a few South Siberian runiform inscriptions, and then it gets generalized from Middle Turkic on. +*gA* must also have characterized Proto-Bulgar and Proto-Oghuz as the dative suffix of the Oghuz languages, of the Volga-Bulgarian epitaphs and of modern Chuvash is +(y)*A*; this would not have been the result of if their original dative suffix would have been +*kA*. +*kA* survives as dative suffix in Khalaj. There is no way to get from +*kA* to +*gA* according to the rules of attested Early Turkic; hence, this must have been a prehistoric isogloss.

The Orkhon inscriptions are written in one of the dialects of their time, politically important but not the only one. They show, among others, the strong consonants /p t k/ alternating with the weak consonants /b d g/, the latter with fricative realizations [β δ γ] and, in suffixes, the vowel archphonemes /A/, /I/ with /i i/,⁸ /U/ and /X/ (= i i ü u and low vowels as well). The archaic features of these inscriptions include several plural suffixes, +*lAr*, +(X)*t* and +*An*, whose use is limited to nouns denoting important persons and relatives; there is here no plural category for non-human entities, nor is +*lAr* added to verb forms. Other archaisms are the retention of the /ñ/ phoneme and of the comitative suffix +*IXgU*, which survives in the language of the Yakut (Sakha), thought to be descendants of the Kurikan. The anticipating counting system of the inscriptions survives in Western Yugur (Yellow Uyghur). The assertive interrogative particle *gU* survives in Modern Uyghur. The inscriptional negative participle suffix -*mAčI* is also found in some archaic Old Uyghur texts. *kām* 'who' as

⁵ This would give the Turkic group a time depth similar to the Germanic sub-family of Indo-European, one much shorter than that of the Iranian languages.

⁶ This name is inherited both by the Volga river (on the banks of which there was the city of Bulgar, the capital of the Volga Bulgarians) and by the country of Bulgaria (because it continues the Danube Bulgarian state, which became Slavic after a few generations because most of its population was Slavic).

⁷ Old Turkic *yog*, 'funeral feast' already appears as *δόγια* in the account of an ambassador of the Byzantine emperor Menander Protector in 576 A.D. (with a Greek suffix), but the term could have been borrowed from a predecessor language.

⁸ The perfect participle and evidential suffix -*miş* and the 3rd person possessive suffix +(s)*i(n+)* may have had non-harmonious /i/.

against *kim* in Old Uyghur, *ärkli* against Old Uyghur *ärkän*, epistemic *ärinč* against Old Uyghur *ärki*, should be considered dialect features and not archaisms. Old Uyghur is innovative as against the runiform sources in using *kim* as clause subordinator, following Iranian models. On the other hand, a number of features show that Orkhon Turkic is not Proto-Turkic: One of them is the replacement of *tägül* ‘is not’ by normalized *är-mäz* in the inscriptions as well as in Uyghur and Khâqânî Old Turkic; it survives in Oghuz Turkic, as Maḥmūd pointed out, and in western Kipchak languages. Onset /h/ survived in Modern Khalaj but not in Old Turkic. The vowel length distinction, which survived well in Yakut, Turkmen and Khalaj, may have been lost in runiform Turkic (though /a:/ : /a/ survives in some Xinjiang runiform manuscripts). The vowels in the interrogative particle *mU* of Orkhon and other Old Turkic sources are secondary as against *mI*, its Ottoman form (neither Old nor Ottoman Turkic had fourfold harmony in it). The loss of the *-(X)gII* participle (except in petrified *ärkli*, used as temporal conjunction), the replacement of *yaz-* ‘to write’ by *biti-*, of *buyur-* ‘to command’ by *yarlıka-*, the loss of *ud-* ‘to follow’, the creation of *+IAyU* as a simulative case suffix, of *+tAg* as pronominal case suffix, of *yan* ‘side’ as a postposition with vowel harmony, all show that the dialect of the inscriptions is in some ways less archaic than even some modern Turkic languages.

We mentioned the Bulgar, the ancestors of the Chuvash, who reached the Volga river, as the first group to split away from the other Turks. Their language had some clear distinctive features, the most famous of which are that Common Turkic /z/ there appears as /r/ and common Turkic /ʃ/ as /l/ or as /lj/. Other important features are the loss of /g/ and sometimes of coda /k/, and a change of /d/ to /r/ between vowels. Instead of Old Turkic or Oghuz word onset /y/, Bulgar had /j/, like the various Turkic language groups in Siberia and most Kipchak languages.⁹ The opposition between long and short vowels was preserved in Volga Bulgarian. There was a proximal demonstrative *ku* surviving also in Yellow Uygur. ‘Who’ was *käm* and not *kim*. The Proto-Bulgars must have formed a distinct ethnic and linguistic entity already in their South Siberian homeland, before they left the close-knit Turkic roaming grounds: All the consonant features mentioned characterize Turkic loans in Proto-Mongolic; all the early lexemes shared by Turkic and Mongolic show these Proto-Bulgar features. This means that the Proto-Bulgars were the first Turkic people who influenced the Mongols. In Eastern Europe, Proto-Bulgar tribes in the 5th century A.D. arrived at the north of the Black Sea, where the population at that time may have been mostly Iranian. They moved to the Volga-Kama region only in the 9th century, where two different varieties influenced Hungarian, a Finno-Ugric language, but were themselves influenced by the Finno-Ugric languages spoken there. The Volga-Bulgarian epitaphs of the 13th – 14th century show additional features unusual for Turkic languages, also found in present-day Chuvash: The 3rd person imperative, unusual in the European languages, was replaced by the 2nd person imperative of the causative in *-tUr-*. To translate this process into Turkish, it is as if one replaced the sentence *İşi yapsın* by *İşi yaptır*. When I say *İşi yapsın* to somebody, I make him responsible for the carrying out of the job by a third party; this is in fact what the causative expresses. The Common Turkic conditional suffix *-sA* was used like a *-(X)p* converb, expressing anteriority in our examples. Volga-Bulgarian did not use *+IAr* as plural suffix; I already mentioned that *+IAr* was not used for general plurality even in the Orkhon inscriptions. While the other Turks generalized the use of *+IAr*, Volga Bulgarian made *+sem* into its plural suffix. Bulgar-Chuvash *+sem* does not precede the possessive suffixes, as *+IAr* does in the other Turkic languages, but follows them, perhaps because it was a noun borrowed from Finno-Ugric. Volga Bulgarian and Chuvash have united the dative and the accusative into one case, with a suffix *+A*. The Chuvash personal pronouns are *epě*, *esě*, *epir*, and *esir*, consisting of a deictic element preceding the forms **bi*, **si*, **biz*, and **siz*. In Old Turkic, **bi* and **si* were the stems not only of *bi+z* and *si+z* but also of the accusative, locative and ablative singular *bi+ni* etc.; Chuvash generalized this stem while Common Turkic retained the low vowel of the nominative and ablative stem. In Volga Bulgarian, the ordinal suffix is *+šI* for numerals ending in vowels, *+Im* for numerals ending in consonants; the latter was perhaps borrowed from Iranian. Chuvash joins the two suffixes in its ordinal suffix *+měš*. Another special feature is the negative particle *an* pre-posed before imperative forms, whereas all other Turkic languages adhere to morphological negation for verbs.

⁹ Modern Uyghur onset /y/ is realized as [j] before high front vowels.

We can then reconstruct Proto-Turkic based not only on Old Turkic but also on direct and indirect Bulgar evidence, on what we learn from the modern languages (especially from the aberrant languages Khalaj, Yakut and Chuvash), and on the Turkic contact features adopted in the Mongolic lexicon and morphology. Proto-Turkic had the same consonants as Old Turkic but some Old Turkic onset /y/s came from Proto-Turkic onset * /d/ and * /n/ (which no longer exist in Old Turkic, the latter only in *nä* ‘what’); real Old Turkic onset /y/ may under certain circumstances have been realized as */j/, the voiced counterpart of /č/. Proto-Turkic had onset /h/ (preserved in Khalaj, in Old Turkic realized as a /y/ ~ Ø alternation), which may have come from /p/ at an even earlier stage. Proto-Turkic suffixes had the vowel archiphonemes */A/, */I/, */O/ (which gave Old Turkic /U/) and */U/ (which gave Old Turkic /X/). Among the case suffixes, +čA had primarily local, perlocative meaning. The genitive of the demonstrative pronoun *a*¹⁰ (oblique stem *in+*) was used as 3rd person possessive suffix. Pronouns had vowel stem alternations, low vowels in the nominative, high vowels in the oblique case forms (*käm / kimni, bo / buni, an(lar) / intin, bän / bini*). Nominal sentences may have been negated with an element surviving as *dag* in Argu and in Khalaj. This element may have had a variant **täg*, Common Turkic *tägül* coming from **täg ol*; but Noghay *tuwıl* has back vowels.

Many Proto-Turkic verbal stems ending in consonants in Old Turkic had further vowels at the end of the stem (preserved in loans to Mongolic and in the vowel converb mentioned below). The deontic modal paradigm probably had a form *-(A)llŋ* when the addressee was meant to be involved in the activity, *-(A)llm* when (s)he was not meant to be involved. The indicative and the epistemic modal forms were all infinite: The past tense forms consisted of a verbal noun in *-d* with possessive suffixes, and the Orkhon inscriptions show that the necessitative in *-sXk* also had a subject inflexion with possessive suffixes. Proto-Turkic may (like Proto-Indo-European and Proto-Semitic) not have had a dedicated future form, modal forms with future projection like *-gAy* or imperfect participles like *-dAčI* serving for future reference. The form in *-gAn* probably had habitative,¹¹ *-mlš* probably resultative meaning (since both the evidential and the perfect depend on evidence of previous events). Such forms, and the imperfect one in *-Ur* (apparently derived from the vowel converb) or the prospective one in *-gU*, served both as verbal nouns and as participles having subject, object or other roles in the event referred to; they could be used for nominal reference, could qualify other nouns or could serve as predicates. Other suffixes marking non-finite verb forms were perfective *-dOk* and *-yOk* and imperfective *-(X)gll* and *-(X)gmA*. There were linking converbs formed with *-(X)p*, *-(X)yXn*, and *-tl*; the vowel converb, also only serving clause-linking, may have been the bare stem (the shorter 2. person singular imperative having gotten syncopated from the stem through onset stress). There also were semantic converb suffixes like *-gAll* and the conditional *-sAr*.

4. Deriving the modern languages from Proto-Turkic

We can now explore some of the ways in which language contacts altered the sounds and the morphology of Proto-Turkic, to result in the systems we have today. One should distinguish Turkic languages in which vowel labialization is consistent throughout the word, like, on the one hand, Western Oghuz, where it applies only to high vowels and Yakut, Kirghiz, and Altay Turkic, where it applies also for low vowels and, on the other hand, language like Kazakh or Turkmen, where vowel labialization tends to get lax and disappears towards the end of long words. Most Turkic languages now have long vowels in loans, but languages like Kirghiz also have secondary vowel length in Turkic stems and suffixes. Turkish has joined the archiphonemes /X/, /I/ and U/ to one fourfold alternation but Modern Uyghur, e.g., keeps them apart. Some languages readily adopt foreign consonants like /h/ or /f/ while others, like Yakut, fully adapt consonants in borrowed items to the Turkic set or change some of them (e.g. /f/ > /p/) to the usual ones in the spoken or even in the written language.

¹⁰ This explains why this suffix showed the pronominal /n/. The Yakut 3rd person possessive suffix *+(t)A(n+)* survives from this stem. *şıgar* ‘towards; half’ may be the dative form of a pronoun which became the possessive suffix variant after vowels.

¹¹ Like Turkish *çalış-kan*, ‘diligent’; cf. *äsnägän bars* ‘a yawning tiger’ in the Irq Bitig, one of the few instances of this form in early Old Turkic.

The adoption of the gender category is very rare (found e.g. in Karaim). Languages dropped some case suffixes but formed some new ones: directive suffix +*gArU* given up, replaced by elements like +*sAr* (< *siŋar*) in some languages, instrumental and comitative forms in +*In* and +*IXgU* given up but replaced by +*bA*, +(y)*IA*, +*nAn* and other forms coming from *birlä*. Some secondary case suffixes were created already in the Old Turkic varieties. The demonstrative system was enriched in most Turkic languages, often by fusion with presentative elements such as *oŝ* or *ha*. Apophony in pronominal inflexion was given up, and analogy from nominal morphology changed pronoun morphology. Most languages introduced the use of 2nd person plural pronouns for politeness. There was wide-spread replacement of the reflexive pronoun *käntü* by the nouns *öz*, *bod*, and *beye*.

The nominal plural suffix +*IAr* replaced some verbal plural markers. Most Turkic languages grammaticalized numerous verbs for expressing forms of actionality, i.e. for describing how actions are carried out. As a further step in grammaticalization, such elements became present tense suffixes, e.g. -*ča* or -*iwat* << *yat*-, -*du* or even -*t* << *tur*-, -*yor* or even -*yAr* << *yor*-. In many languages, personal pronouns cliticized or became suffixes to form present tense inflecting verb forms. In a parallel process, Turkic languages except Yakut dropped the /r/ of the conditional converb (which did not inflect for person, though it could, of course, be accompanied by pronouns referring to the subject) and received possessive person inflection like the past tense form – no doubt through contact with the early Indo-European and Uralic languages which have finite conditional forms. Only Yakut, which only had contact with Mongolic and Tungusic languages, preserved the conditional form practically unchanged. A number of participle/verbal noun forms were lost, e.g. in the one -(X) *gII*, or survived only in marginal Turkic languages, e.g. the one in -*dOk*; but new ones came up, either from the upgrade of derivational suffixes, e.g. the one in -(X)š (> -(y)Iš), or by morphological fusion, e.g. -*idiyan* < -*a turgan*, or through splits, e.g. -*gAn* giving the suffixes -*An* beside -*yAn* in Turkmen, or with the addition of particles, e.g. -*AčAk* < -*gA čak*, or with the addition of derivational suffixes, e.g. Turkmen -*mAll* < -*mA+IXg*. In Proto-Turkic, ability and inability were expressed by the verb *u*-. morphological form expressing inability was formed from this in Old Turkic but survived only in Western Oghuz; other Turkic languages came to use *al*- ‘to take’, *bil*- ‘to know’ and *bol*- ‘to become’ as auxiliaries expressing this content, with *al*- becoming a suffix -*Al*- in some of them. In synthetic relative clauses qualifying or representing non-subject entities involved in the event, reference to the head is the task of independent pronouns, of possessive suffixes added to the head or added to the infinite verb forms. New converb forms were also created, e.g. by adding the derivational suffix / particle *rAk* to give -(y)*ArAk*.

5. Summer camps and winter camps¹²

The Kazakh speaker who addressed the ConCALL conference just before me mentioned the words *jaylaw* and *kıstaw*, respectively denoting the ‘summer camp’ and the ‘winter camp’ of the Kazakh semi-nomads. A Persian participant at the conference, an Iranist, then remembered the terms *yayla:q* ‘summer camp’ and *qišla:q* ‘winter camp’, used in earlier Persian texts describing the life of non-sedentary groups. The Persian terms must be loans from Turkic, since they are analyzable in Turkic but not in Persian. We will here follow these terms in the Turkic languages to illustrate their development. The Proto-Turks were semi-nomads migrating between summer and winter camps and must have pronounced these two words as *ya:y+lag* and *kiš+lag*. Other early +*IAg* terms denoting terrain are *tarig+lag* ‘arable field’, *r+lag* ‘territory with wild vegetation’ and *turug+lag* ‘place of residence’. In Old- and thence in Proto-Turkic, ‘spring’ appears to have been *yaz*, ‘summer’ *yay*, and *kiš* ‘winter’.¹³ Turkish (like a few other Turkic languages) switched to the Persian

¹² Many of the terms discussed in this section are listed in §1941 and §1496 of Gerhard Doerfer, *Türkische und Mongolische Elemente im Neupersischen*. Wiesbaden, Steiner, 1963-1975, with loans mentioned as well.

¹³ Not all of Old Turkic may have preserved the vowel length in *ya:y*, presumably because its direct contact languages did not have a distinction between ‘normal’ and long vowels. The confusion between *yaz* and *yay* is documented in Gerard Clauson’s *Etymological Dictionary of Pre-thirteenth Century Turkish* (Oxford, Clarendon 1972), but that confusion is not really important for our purposes.

term *bahar* (or *ilkbahar*) for ‘spring’ and moved *yaz* to ‘summer’, so that Turkish *yayla* (< *ya:ylag*) is no longer transparent. It now signifies ‘plateau’: a metonymic shift, since plateaus served for summer pasture. We see a different metonymic shift in Turkish *kışla* (< *kışlag*), which denotes ‘military barracks’: Soldiers need to live in barracks in winter but not in summer. Both semantic shifts show abandonment of the nomadic life style, the former through a widening, the latter through a narrowing in meaning. Turkish also has *kışlak* with the old meaning, ‘nomadic winter camp’, but that is a loan either from Persian or from Chaghatay Turkic, the source from which Persian took the word.¹⁴ This term shows a split survival also in Uzbek, which has *qişlâv* ‘winter camp of the nomads’ vs. *qişlâq* ‘village’: In this case the form with the regular sound change retained the original meaning while the form inherited from literary Chaghatay Turkic, the language of the Timurid empire, carries the semantic shift away from nomadism. The Uzbek were Kipchak nomads who vanquished and conquered the Chaghatay state but adopted its culture; this fusion is shown in these two terms. Chaghatay Turkic was the written standard for all Muslim Turkic peoples outside the Ottoman empire, between Beijing and Kiev; *qişlaq* and *yaylaq* were therefore adopted also by a number of Iranian languages and dialects, *qişlaq* also by Russian, Byelorussian and Ukrainian. On the other hand, Ottoman *qışla* found its way into Arabic dialects and into Balkan languages, whose speakers were ruled by the Ottomans.

Our main use for these two words will be to serve as examples for the action of sound laws, and first the vowels. The first syllable long vowel of *ya:ylag* is retained in Turkmen *ya:yla*, which also kept the meaning ‘summer abode’: Most Turkmen were nomadic till the early 20th century. Uzbek and Modern Uyghur cancelled the opposition between /i/ and /ɪ/; this already happened in their predecessor language Chaghatay, under the influence of both Persian and Mongolic. We already mentioned the Uzbek forms *qişlâv* and *qişlâq*, and Modern Uyghur has also inherited Chaghatay *qişlaq*.

Kazakh and South Siberian Khakas have changed /š/ to /s/, thence the /s/ in Kazakh *kıstaw* and *qıstağ* in the Khakas dialects. I take š > s to be the result of Mongolic influence, as š did not exist as a phoneme in Classical Mongolian, but only as an allophone of /s/ appearing before /i/.

The Oghuz languages dropped all coda /g/s, whence Turkmen *ya:yla* and Turkish *yayla*. Under Mongolian influence, Chaghatay coda /g/ became /k/, realized as [q] in back-vowel words; we already mentioned the Chaghatay terms ending in *q* retained in Modern Uyghur and Uzbek and borrowed into Persian and from there into Ottoman and Modern Turkish. Khalaj, which has *qişlay* / *qişla:γ*, retained coda /g/. So did the Yenisey and Sayan groups of Siberia: We have Khakas *čaylağ* and *qıstağ* and Tuvan *čaylağ* and *qıstağ*. On the other hand, a 14th century Kipchak dictionary already has the variant *yaylaw*, showing the Kipchak labialization of /g/ after low vowels which we see e.g. in *tag* ‘mountain’ > *taw*.¹⁵ We started out with Kazakh *ǰaylaw*, already mentioned Uzbek *qişlâv*. and the West Siberian Baraba Tatar dialect shows *qişlaw*. In Kirghiz, this sound sequence *aw* was regularly changed to *o:*, e.g. in *to:* ‘mountain’; thus we have Kirghiz *ǰaylo:* and *qišto:*. Teleut, a dialect close to Altay Turkic, further changed this *o:* to *u:*; so this word there changed to *yaylu:*.

Proto-Turkic onset *y-* stayed unchanged in Oghuz, Khalaj and Uzbek; thus the ‘summer camp’ word begins with *y-* in these languages. This is also the case in most of Western Kipchak, i.e. in Karaim, Krymchak, Crimean Tatar, as well as in Kumyk, Bashkir and a part of the Volga-Kama Tatar dialects (e.g. Misher Tatar). Other Kipchak languages and varieties changed onset /y/ to /j/, as already mentioned by Maḥmūd in the 11th

¹⁴ Other examples for Persian loans into Turkish retaining the original Old Turkic sound shape are Turkish *kadın* ‘woman’ vs. borrowed *hatun* ‘lady’, and Turkish *oda* ‘room’ vs. borrowed *otağ* ‘large, ceremonial tent’. These lexemes originally come from *xa:tun* and *o:t+a-g* respectively; in both words the /t/ became /d/ because it follows a Proto-Turkic long vowel and the Oghuz drop of the coda /g/ in *otag* is also regular.

¹⁵ This very early Kipchak process is seen also in Bulgar-Chuvash and in South Siberian Altay Turkic. Crimean Tatar and Krymchak, the Jewish dialect of Crimean Tatar, are Kipchak languages in which it did not take place, presumably under Ottoman influence; it is not consistent in Siberian Tatar, presumably under the influence of neighboring Yenisey Turkic. The Kipchak words for ‘island’, *ataw* (< **a:tag*; cf. Turkish *ada*) and *otraw* / *utraw* (< **otrog*), show the same process.

century;¹⁶ thus we saw Kazakh *ǰaylaw* and Kirghiz *ǰaylo*:. In South-Siberian and also Kirghiz-Kipchak Altay Turkic, *y*- changed to palatalized dental *dʹ-/tʹ*-, in other South Siberian languages to *č*- (whence Khakas and Tuvan *čaylaǰ*), in Yakut then further to *s*-.

Finally, some of these languages have varying types of cluster assimilations, for which we here have the evidence of Shor and Tuvan *qıǰtaǰ*, Kirghiz *qıǰto*:, Kazakh *kıstaw*, and Khakas *qıstaǰ*.

6. Conclusion

We see that a non-systematic and partial documentation of the evidence for only two words can show how accountable and regular sound processes are able to explain all the Turkic varieties. A short paper cannot, of course, detail all the processes that led to the emergence of all 40 modern languages; but I hope to have shown that – in the domains of phonology, morphology, syntax and semantics – this task is feasible and that it would be a highly useful exercise not only for understanding the history of the Turkic languages, but also for helping speakers of one Turkic language to switch to other Turkic languages both as to their passive and their active abilities.

¹⁶ *y*- > *ǰ*- was a prehistoric process in Bulgar Turkic, as shown by Mongolic cognates, and we still find *ǰ*- in 14th century Volga Bulgarian; in modern Chuvash, however, words with original *y*- > *ǰ*- changed to *ś*-, a palatal sibilant.

Lessons Learned from the Study of Mongol Languages

György Kara
Indiana University

Abstract

After more than two centuries of increasingly intensive research on these languages, the field still offers a great many of problems to be solved and sources for building or testing theories. These languages that form a small family show manifold outer contacts and far-reaching influence. The issues to be discussed include typology (CSCP or “left-branching” syntax, mostly “right-branching” agglutination; vowel harmony, phonotactics); disputed classifications of the living languages and dialects; periodization; “Mongolic” and “Para-Mongolic”; Mongol influence on other languages; influence of other languages on Mongol; two millennia of the Altaic Sprachbund; questions to be answered, fields to be explored.

A stylized map of a region, possibly a country or a large administrative area, divided into several colored regions. The colors include orange, blue, yellow, brown, and green. The map is tilted slightly to the right. The text "Part II: Formal Linguistics and Research" is overlaid on the map.

Part II: Formal Linguistics and Research

An Acoustic Analysis of Vowel Insertion at Syllable Edges in Turkish

Hassan Bokhari and Mustafa Durmaz
Indiana University

Jonathan Washington
Swathmore College

Abstract

Two distinct vowel insertion processes have been identified in Turkish: one that splits underlying onset clusters, and one that splits underlying coda clusters. Existing literature suggests that the two processes differ in consistency of application, quality of the inserted vowel, and whether these differences are subject to register effects. Bellik (2016a, 2016b, 2016c) proposes a grammatical difference between the two types of insertion in Turkish that captures their different behavior. Vowel insertion between coda clusters is claimed to be phonological epenthesis, while insertion between onset clusters is argued to be phonetic intrusion. Epenthesis and intrusion are argued to differ in terms of phonological presence of the vowel and specification of a gestural target.

We investigate whether there is a detectable difference between vowels inserted to split underlying onset clusters and underlying coda clusters by examining the phonetic properties of the vowels. Data was collected from four native speakers of Turkish in the form of a list-reading task, with target words embedded in carrier sentences. We measure the frequency of vowel insertion in both environments, the formants of the inserted vowels, and the duration of the vowels, and compare these measures with those of underlying vowels in the same conditions.

Our findings confirm that vowels are not consistently inserted in onset clusters, while they are entirely consistent in coda clusters. In addition, vowels inserted in onset clusters have schwa-like formant values and their durations are overall shorter than coda-inserted vowels and underlying vowels. Vowels inserted in coda clusters are nearly identical to underlying vowels in terms of length and formant values. All of these pieces of evidence lead us to agree with Bellik (2016), in that vowels inserted to break onset clusters in Turkish are phonetically intrusive and lack a gestural target, while vowels inserted in coda clusters are phonologically epenthetic.

Keywords: phonetic analysis of vowels, vowel insertion, Turkish language

1. Introduction

In Turkish there are two processes of vowel insertion: one that splits underlying onset clusters, and one that splits underlying coda clusters (Göksel & Kerslake, 2005, 12-13, 17-18; Clements and Sezer, 1982, 243-248). Processes of vowel insertion between two consonants in a syllable margin are attested in a number of other languages (Hall, 2011).

Vowel insertion between onset clusters in Turkish usually occurs in words where the vowels are not written, as in “tren” [tiren], “plan” [pulan], “spor” [sipur] or [ispor]. Words with onset clusters are all borrowings, mostly from European languages.

Turkish allows coda clusters with falling sonority to surface (Kornfilt, 1997, §3.2.2; Göksel & Kerslake, 2005, p. 13; Topbaş and Kopkallı-Yavuz, 2008), but only when the second consonant is an obstruent (Levi, 2001, p. 384), for example [tejp] ‘tape recorder’, [ilk] ‘first’, [renk] ‘color’, [ʃevk] ‘enthusiasm’. All other types of underlying coda clusters receive an inserted vowel if the two consonants cannot resyllabify into separate

syllables (Kornfilt, 1997, §3.2.6; Göksel & Kerslake, 2005, p. 17), such as /karn/ [karun] ‘belly’, /ʃehr/ [ʃehir] ‘city’, and /œmr/ [œmyr] ‘lifetime’, which alternates with forms where insertion does not happen, e.g. /karn-u/ [karnu] ‘his/her belly’, /ʃehr-i/ [ʃehri] ‘his/her city’, /œmr-y/ [œmyr] ‘his/her lifetime’. Note that when a high vowel is underlying between two consonants at the end of a word, it does not delete in these contexts, e.g., /kœmyr/ [kœmyr] ‘coal’, /kœmyr-y/ [kœmyry] ‘his/her coal’, reaffirming that this is a process of insertion and not deletion. Many words with underlying rising-sonority coda clusters are borrowings from Arabic and Persian.¹

In general, vowels inserted in both onset clusters and coda clusters are high vowels that harmonize with an adjacent vowel—that is, they take their backness and rounding specifications from them. However, Clements and Sezer (1982, 243-244, 247) show that quite a few onset and coda clusters can take high vowels which are not predicted by harmony. For example, /kabr/ [kabir] ‘grave’ has a front vowel inserted when a back one is expected, and /kredi/ [kuredi] ‘credit’ has a back vowel inserted when a front vowel is expected. However, Bellik (2016a, 2016b, 2016c) claims that the actual quality of the onset-inserted vowels may not match previous descriptions of their quality.

Vowels inserted in onset clusters are not written, are potentially very short, do not occur in some registers, and vary in these ways from speaker to speaker. Clements and Sezer (1982, p.246) claim that the inserted vowel is pronounced in “normal or colloquial pronunciations”, while it may not be pronounced in “careful or learned pronunciation”. On the other hand, vowels inserted in coda clusters are pronounced as full vowels, are always written, are not subject to register differences, and are not reported to vary from speaker to speaker. Both types of inserted vowels usually harmonize in backness and rounding to an adjacent vowel (with exceptions, as already noted).

On-going work by Bellik (2016a, 2016b, 2016c) proposes that there’s a formal difference between the two types of insertion. Vowel insertion between coda clusters is claimed to be epenthesis at the phonological level. Evidence for this includes its agreement in backness and rounding harmony, and the consistency of its occurrence. In contrast, insertion between onset clusters is argued to be phonetic intrusion per Hall (2006), who argues that phonetically intrusive vowels have no phonological presence and lack a gestural target. Because of this, an intrusive vowel in a word like /kredi/ should take on the quality predicted by the position of the tongue based on the surrounding context, and not necessarily agree with surrounding vowels in backness and rounding. To determine the difference between onset insertion and coda insertion, we investigate whether these two types of inserted vowels are phonetically similar or not, by measuring frequency of pronunciation, vowel formants, and duration.

The main question of this study is whether vowels inserted in onset clusters and vowels inserted in coda clusters are similar in terms of frequency of insertion, and quality and length of the inserted vowel. If there is a difference in frequency of insertion, vowel duration, and/or vowel quality, and the vowels inserted in onsets are less frequent, shorter, and/or centralized, then this is good evidence to support Bellik’s (2016a, 2016b, 2016c) proposal that vowels inserted to break onset clusters in Turkish are intrusive and occur at the phonetic level, while vowels inserted to break coda clusters in Turkish are epenthetic and occur at the phonological level. Specifically, based on sources like Clements and Sezer (1982) and our own knowledge of the language, we expect the rate of insertion in coda clusters to be near 100%, and the rate of insertion in onset clusters to be a lot lower, especially for speakers of Turkish who know English as well. In onset clusters, if the inserted vowels are indeed gestureless, we should expect a range of vowel qualities due to influence from surrounding consonants, as well as an overall more neutral or schwa-like quality and shorter duration, differing from vowels with phonological targets. Vowels inserted in coda clusters, on the other hand, should have much more stable vowel qualities and longer durations that resemble those of underlying vowels.

¹ Note that the fact that inserted vowels in both onset and coda positions in Turkish appear as the result of an insertion process is not a given: Kabak (2003, p. 57) argues that inserted vowels in Korean words of English origin—such as [pasu] ‘bus’—are actually underlying.

2. Methods

2.1 Participants

The participants in this study all identified as speakers of standard Turkish, regardless of their region of origin, and at the time of the study, they all resided in Bloomington, Indiana, and were aged 20 to 35. Hence, knowledge of English and age group were roughly equivalent. We recorded four participants, and balanced the number of male and female participants (two each). Additionally, we asked how long each participant has lived in the U.S. in order to ensure that level of knowledge of English was mostly equivalent between the speakers. There was, however, a wide range, from 6 months to 4 years, with an average of about 1½ years.

2.2 Stimuli

The participants were given sentences in Turkish that contain words with underlying onset and coda clusters but that normally have inserted vowels between these, as well as a set of words that contrast with these words in that they have underlying vowels in the same phonological environment.

The underlying consonant clusters in the stimuli all consist of stop + liquid: /br/, /kl/, /kr/, /dr/. In order to compare the quality of underlying and inserted vowels, all of the chosen words have [i] between the consonants, except possibly between initial /k/ + liquid (Clements and Sezer (1982) assume that [u] is inserted in onset clusters beginning with velars, regardless of the backness of the following vowel). Table 1 illustrates the stimuli that were used, with each stimulus grouped by consonant context (/br/, /kl/, /kr/, /dr/) and onset versus coda occurrence and underlying versus non-underlying vowel status. The four contexts are underlying coda cluster with potentially inserted vowel (CC), underlying onset cluster with potentially inserted vowel ([CC]), and an underlying vowel in the same phonological context as the previous underlying clusters for onset ([CVC]) and coda (CVC).

	BR	KL	KR	DR
CC]	kabir ‘grave’	sekil ‘shape’	fikir ‘opinion’	kadir ‘worth’
[CC	Bret ‘Brad’	klip [video] ‘clip’	krem ‘cream’	drenaj ‘drainage tube’
CVC]	tabir ‘expression’	tekil ‘isolated’	bakir ‘virgin’	sedir ‘a type of couch’
[CVC	birim ‘unit’	kilim ‘rug’	keriç ‘lime (material)’	dirsek ‘elbow’

Table 1: Words used as stimuli, arranged by onset/coda position, underlying/inserted status, and surrounding consonants

A carrier phrase was used so that the words were in a similar prosodic environment. Each word was inserted into the phrase “Hasan _____ kelimesini yazdı.”, meaning “Hasan wrote the word _____.” Each sentence was repeated twice, and the individual sentences were presented in random order with filler sentences about other things Hasan did. There are 7 filler sentences, repeated 6 times each. The Appendix presents all of the sentences with stimuli and the filler sentences. The total number of sentences presented to each participant was 74.

2.3 Procedure

We recorded speakers of Turkish in Bloomington, Indiana in a quiet setting on mobile devices (i.e., smart phones) with an adequate quality of recording for reliable duration and formant measures. The

randomized stimuli were presented on several sheets of paper, in large print. Because Bellik’s (2016a) study manipulates “speech style and word familiarity to determine how these factors affect” when a vowel is inserted to break an onset cluster, the participants were asked to read the sentences in a colloquial or relaxed way in order to elicit a consistent speech style from all participants. Some background information was also collected, including approximate age and how long they have been living in the U.S. This data is presented in the Appendix.

2.4 Analysis

We measured the duration of all inserted vowels, as well as the first and second formants as a measure of vowel quality. Because vowels inserted between onset consonants were predicted to vary from speaker to speaker in terms of frequency of insertion as well—i.e., there will not always be an inserted vowel between onset consonants—we also counted the frequency of whether a vowel was inserted or not. These three main variables (vowel duration, first and second formants, and frequency of insertion) were compared between the various conditions.

In addition to the inserted and underlying vowels in the stimuli, most of which should be [i]-like, we also measured vowels from non-target words in order to establish reference points for formant measures. These consisted of 8 instances each of the first /e/ in “kelimesini” and the last vowel /u/ in “yazdı”, taken from random instances of these words throughout the recordings, but using the same instances for each speaker. The first of these vowels should vary between cardinal vowels 2 [e] and 3 [ɛ], depending on the speaker’s phonological and phonetic treatment of this vowel, and the second should be approximately cardinal vowel 16 [u]. Measurements of [e] provide a vowel of a different height than [i], and measurements of [u] provide a vowel of a different backness than [i]. This helps establish a rough idea of the speaker’s vowel space in order to determine the relative position in the vowel space of the vowels in the stimuli.

The target words and control words were all marked on a single tier of a TextGrid in Praat (Boersma & Weenink, 2016). Then each vowel to be measured was marked on a second tier of the same TextGrid, and labeled by its cluster condition (e.g., “krem” on tier one marking the interval containing the whole word, and “[CC” on tier two marking the interval for the inserted vowel). The screenshot in Figure 1 provides an example of this for the sentence “Hasan tabir kelimesini yazdı”.

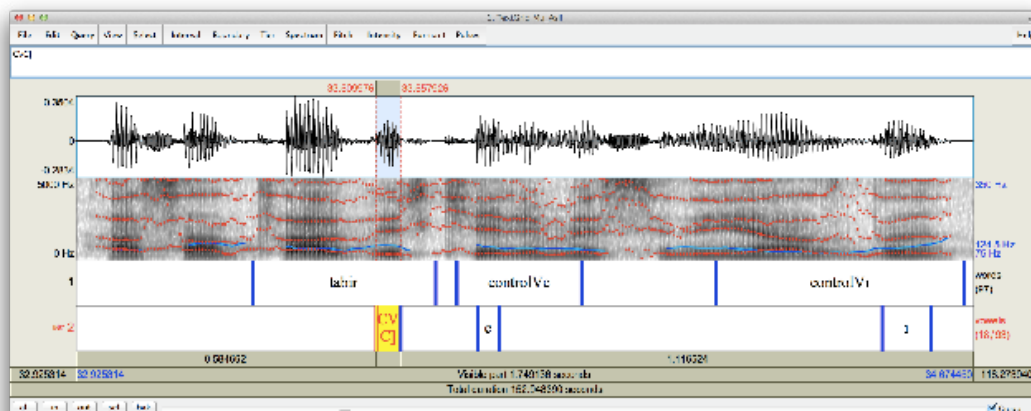


Figure 1: Praat TextGrid demonstrating the two annotated tiers

The edges of each vowel were determined by using cues from the waveform, spectrogram, and auditory judgements, and an attempt was made to select zero crossings in the waveform as boundaries. The screenshot in Figure 2 demonstrates the selection of one of the target vowels.

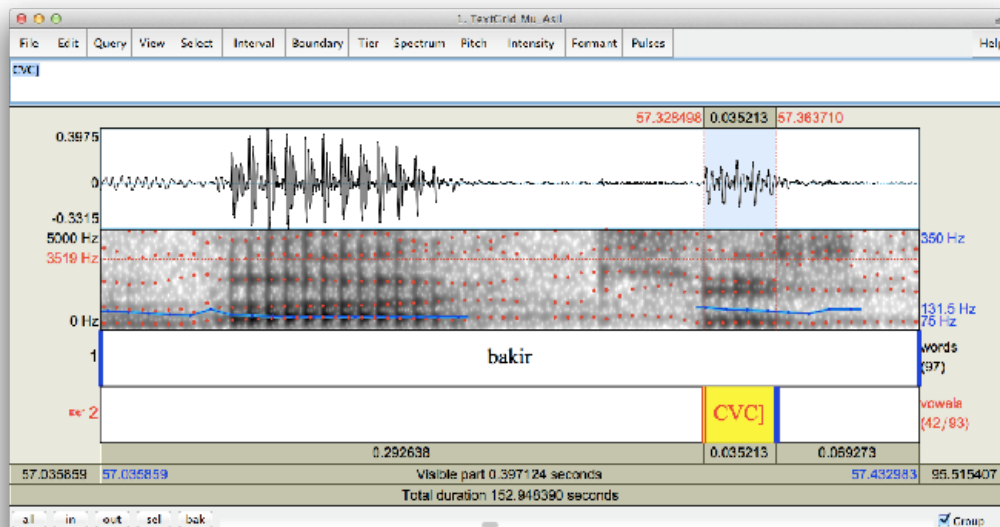


Figure 2: Praat TextGrid demonstrating the alignment of a vowel

In cases where a vowel was not pronounced, nothing was added to tier two. The division between the vowel and a following [r] was sometimes not clear because in Turkish [r] is sometimes realized as a rhotic approximant or fricative instead of as a flap (especially by female participants), but using the waveform, spectrogram, and auditory judgments, it was always possible to find an appropriate edge. Surprisingly, the edges of the [l]s were easier to identify, because much of the time, the [l]s were pronounced as lateral fricatives (especially for female participants). A Praat script was used which extracted for each target vowel the target word, the consonantal context, the duration, and the first and second formants at the midpoints into a text file.

An R script was written which read the text file output by the Praat script and plotted the formants of the target vowels by consonantal context, as well as the reference vowels. Each context and reference vowel is represented by an oval. The midpoint of the oval is plotted at the mean F1 and F2 for the context, and the radius of the height is the standard deviation of the F1 of the vowel measurements and the radius of the width is the standard deviation of the F2 of the vowel measurements. The labels of each oval (i.e., the context or IPA symbol for the reference vowel) were plotted at the midpoints. Underlying vowels were plotted using dark blue, inserted vowels were plotted using light blue, and reference vowels were plotted using green. These colors were chosen so that patterns could be detected quickly and easily on visual inspection of the plots.

A spreadsheet program was used to plot the durations. The cluster conditions were represented as box plots. The box of each box plot represents the range of half of the data, and the midline of the box represents the median, while the whiskers represent the range of all of the data.

3. Results

This study aims to investigate whether there is any difference between the vowels inserted in underlying onset clusters ([CC] and underlying coda clusters (CC)) and underlying vowels that exist in otherwise identical phonological contexts ([CVC and CVC]). Following Bellik (2016a, 2016b, 2016c), vowels inserted in [CC] contexts are expected to have an inconsistent frequency of occurrence (i.e., not 100%), have a different quality than underlying ([CVC and CVC]) and other inserted (CC) vowels, and have shorter duration than other vowels. This study investigates all three of these predictions. We measure frequency of occurrence of the vowel, vowel quality (F1 and F2), and vowel duration. The results of these measurements are presented below.

3.1 Frequency of pronunciation

As discussed earlier, vowels inserted in onset clusters in Turkish are not written and may not be consistently inserted, and whether they are inserted or not may be dependent on register or education level; however, vowels inserted in coda clusters are pronounced and written consistently. Bellik (2016a, 2016b, 2016c) proposes based on this evidence that vowels inserted in onset clusters and vowels inserted in coda clusters are being treated differently by Turkish speakers. In order to investigate this claim, the number of times a vowel was uttered in each condition was counted, and is presented in the graph in Figure 3 as a percentage of the total number of forms uttered for each condition. The results are organized by condition and participant.

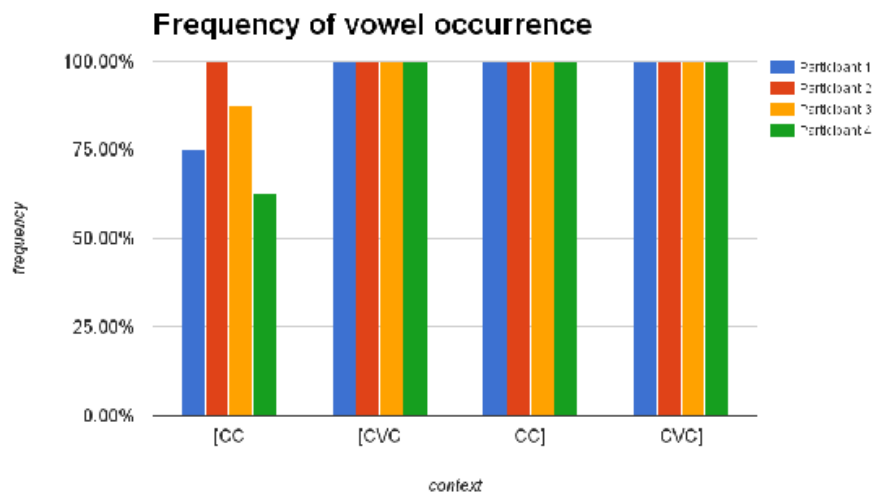


Figure 3: Frequency of vowel occurrence by context and participant

As the graph illustrates, vowels in the [CVC, CC], and [CVC] contexts are all pronounced 100% of the time by all of the participants. In regard to the [CC condition, there is variation. The first participant pronounces the vowel in [CC 75% of the time (i.e., 6 of 8 tokens), the third participant pronounces the vowel 87.5% of the time (7 of 8 tokens), and the fourth participant pronounced the least number of vowels in this condition, at 62.5% of the time (5 of 8 tokens). The second participant pronounced the vowel in this condition 100% of the time. The words where the vowel was not uttered were “klip” (3 times), “krem” (2 times), and “Bret” (1 time). It seems that word frequency may have an effect on whether a vowel is pronounced in this condition, since the most common words are most likely to be pronounced without vowel insertion in the onset cluster, and the least common word (“drenaj”) always received vowel insertion. However, there is not enough data in this study to be sure about the existence of such a frequency effect.

These findings confirm that vowel insertion in [CC] condition is more stable than in [CC, which supports Bellik’s (2016a, 2016b, 2016c) and Clements and Sezer’s (1982) claims about variation in this condition.

3.2 Vowel Quality

Bellik’s (2016a, 2016b, 2016c) claim that vowels inserted in [CC conditions differ from other vowels in their lack of an articulatory gestural target predicts that their quality is not stable and is affected by the surrounding consonants, or is like a schwa vowel. The graphs in Figure 4 present, for each participant, measurements of F1 and F2 for the vowels in the [CC condition relative to the vowels inserted in [CC] condition, the underlying vowels in [CVC and CVC] conditions, and two reference vowels, [e] and [u].

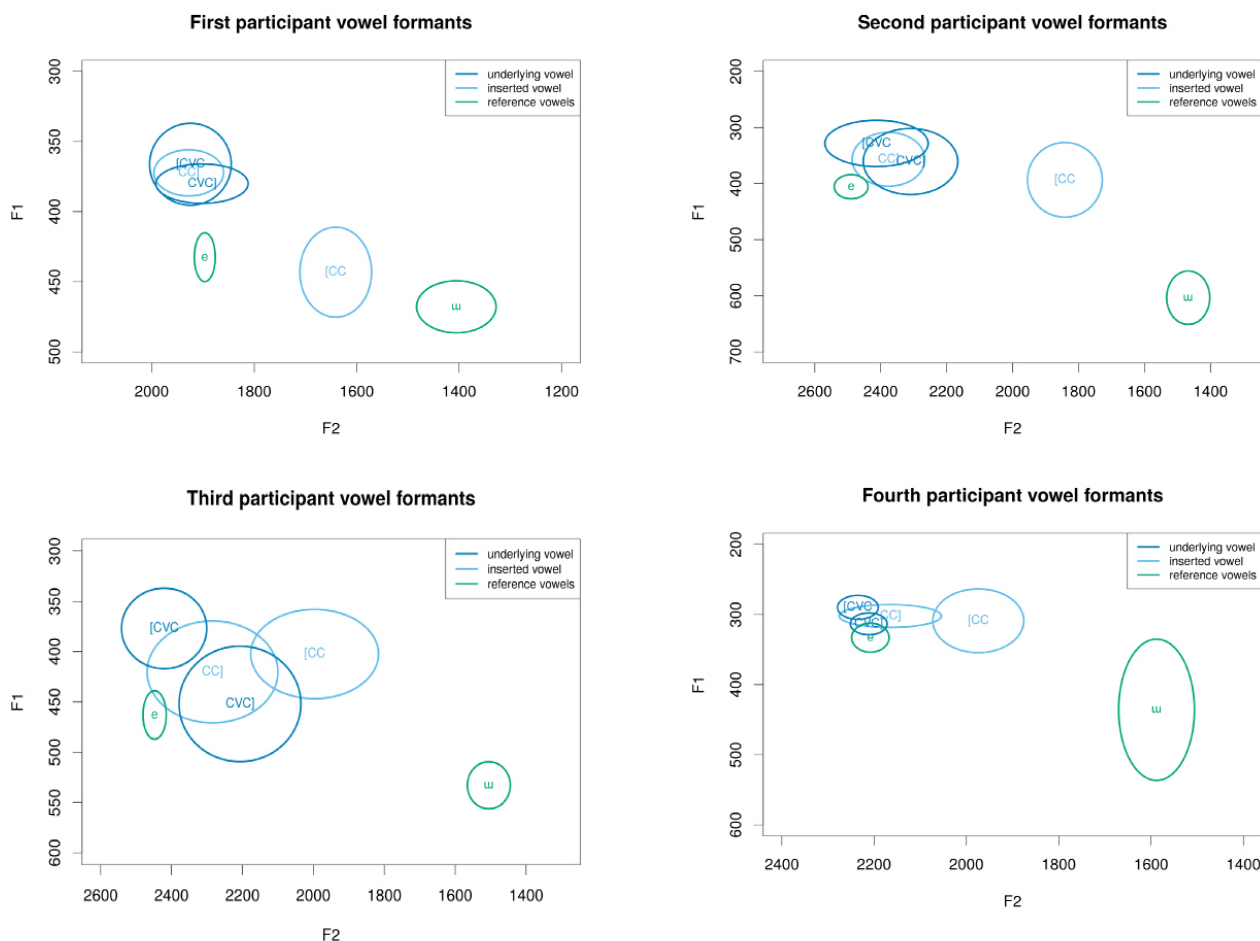


Figure 4: Vowel formants for underlying, inserted, and reference vowels for each of the 4 participants

As the graphs illustrate, vowels in [CVC and CVC] and CC] conditions all overlap with each other in both F1 and F2 for all speakers. The F1 for these vowels is centered around 325-375Hz for men and 350Hz-425Hz for women, and the F2 for these vowels is centered around 1950-2250Hz for men and 2300-2400Hz for women. The average formant frequencies for American English [i] are reported to be 342Hz (F1) and 2322Hz (F2) for men and 437Hz (F1) and 2761Hz (F2) for women, and the average formant frequencies for [u] are 427Hz (F1) and 2034Hz (F2) for men and 483Hz (F1) and 2365Hz (F2) for women (Hillenbrand et al., 1995, p. 3103). For the Turkish speakers in the present study, F1 is low, closer to American English [i], while F2 is low, closer to American English [u].

The vowels inserted in [CC condition for the first two speakers have entirely distinct first and second formants from all other vowels examined, while the vowels inserted in [CC condition for the third and fourth participant overlap some with other non-reference vowels, especially in CC] condition. The F1 value of the vowel in [CC conditions for the second, third, and fourth participants is close to the F1 value of the [i] vowel in the other non-reference conditions. For the first participant, the F1 value of the vowel in [CC conditions is similar to the reference [u] back vowel condition. F2 for the vowel in [CC conditions for all participants is between the F2 values of reference [u] and the other conditions which are assumed to be [i]. In summary, for all speakers the vowel in CC] context is more back than other [i] vowels, and sometimes lower as well. It's also more front, and somewhat higher than the reference [u] vowel. The evidence suggests that underlying vowels and the vowel inserted in CC] conditions are like [i]. The vowel inserted in [CC conditions, however, is much more like schwa.

3.3 Vowel Duration

If vowels inserted in one of the conditions are not specified phonologically, this predicts that they will be shorter than vowels in other conditions. The graphs in Figure 5 present the duration measurements for each of the conditions for each participant.

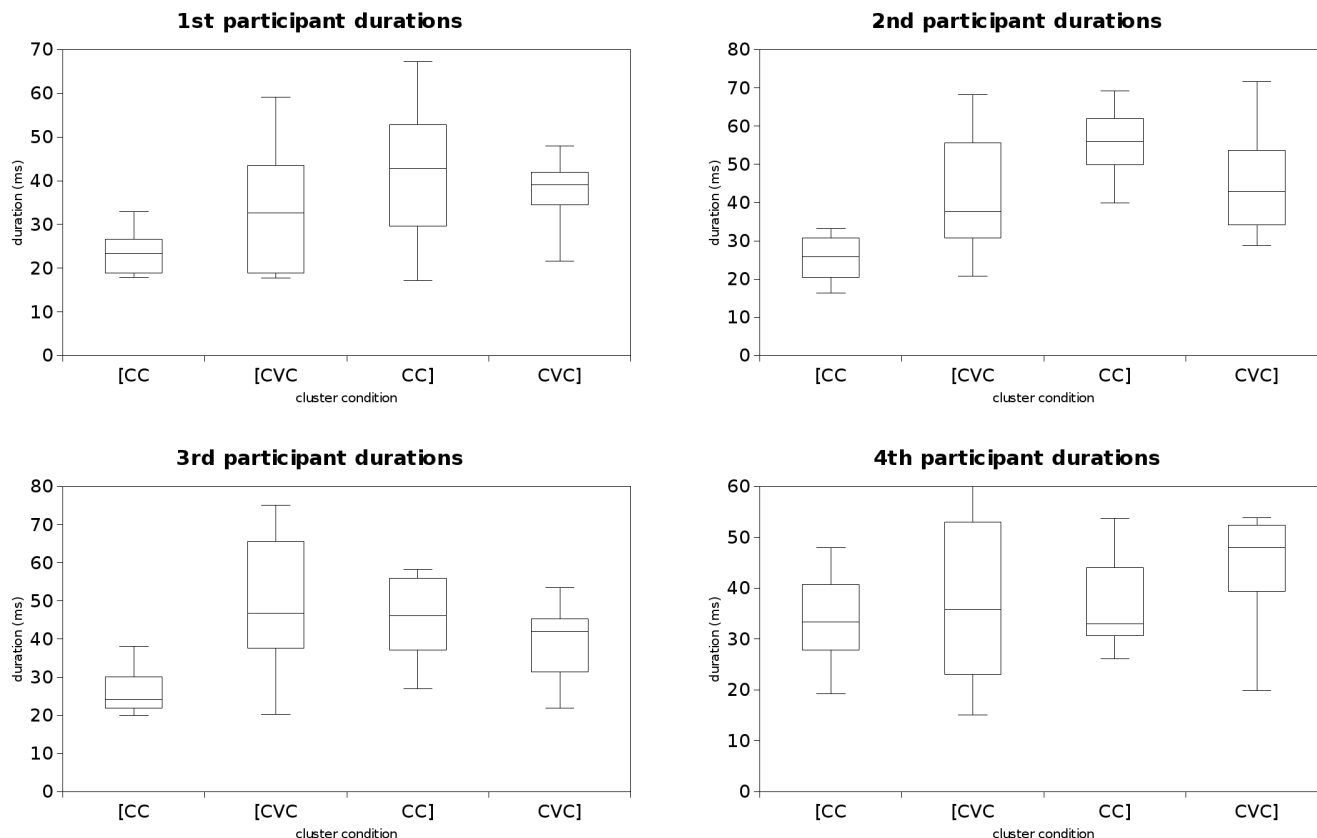


Figure 5: Vowel durations in each context for each participant

The duration measures for the target vowels for each of the first three participants show that the durations of vowels in [CVC, CVC] and CC] contexts have similar ranges as one another, while the durations of vowels in the [CC context have a much smaller range, and the top of range is lower than the median for the vowels in the other conditions. For the fourth participant, the range of durations in all conditions is similar, but the top of the range of [CC vowels is somewhat less than for the other conditions. The shorter duration of vowels inserted in [CC conditions supports the idea of inserted vowels in these conditions being unspecified for a phonological or articulatory target, since such vowels (i.e., schwa-like vowels) are expected to be shorter.

4. Discussion and Conclusion

Following Bellik (2016a, 2016b, 2016c), we hypothesized in this study that vowels inserted between onset consonants ([CC condition) in Turkish are different from vowels inserted in coda clusters (CC]) and underlying vowels in similar conditions ([CVC and CVC]) in terms of frequency of pronunciation, quality (F1 and F2), and duration. Specifically, we expected that vowels inserted in codas and underlying vowels would be pronounced 100% of the time, while vowels in [CC condition would be pronounced less often, and would vary from speaker to speaker. In addition to that, the quality of vowels inserted between onset consonants was expected to differ from the vowels in the other conditions, and the duration of vowels in [CC condition was expected to be lower than in the other conditions. All of these expectations were met, confirming our hypothesis.

In this study, we measured the frequency of vowel pronunciation, the quality of the vowels (F1 and F2), and the duration of these vowels in the following conditions: [CC (vowels inserted between onset consonants), CC] (vowels inserted between coda consonants), and [CVC and CVC] (underlying vowels in similar contexts). If the frequency of pronunciation of the vowel in [CC condition is lower than for other vowels or varies, the quality is significantly different, and the duration is lower, then Bellik's (2016a, 2016b, 2016c) hypothesis that vowels inserted in [CC contexts in Turkish are phonetically intrusive, while vowels inserted in CC] contexts are phonologically epenthesized can be supported. Our findings support this claim.

This study finds, for two male and two female speakers of Turkish, that vowels in [CC conditions have variable frequency of pronunciation, a schwa-like quality, and shorter durations than vowels in other conditions. Vowels in other conditions are pronounced 100% of the time, while vowels are inserted between onset clusters around 75% of the time. The F1 of these inserted vowels is like either [i] or [u], depending on the speaker, but its F2 is always between that of these two vowels. The range of duration is always less than vowels in other contexts, and the longest of the vowels inserted between onset consonants for most speakers is shorter than the average duration of vowels in other conditions. The variation in frequency of use of the vowel in [CC conditions, its schwa-like quality, and its short duration all support the view that it has no articulatory target. We further conclude that in [CC conditions in Turkish, this intrusive vowel is not phonologically specified (but the epenthetic vowel in CC] is) and is only uttered as phonetically necessary, meaning that it is an intrusive vowel on the level of phonetics, and not an epenthetic vowel on the level of phonology. These acoustic findings corroborate the articulatory findings of Bellik (2016a, 2016b, 2016c). It is unclear how this distinction might be born out in perception, but Sezer (2016) represents on-going work on the perception of inserted vowels in Turkish onset clusters.

In future work, we plan to investigate vowels in a wider range of consonantal contexts, although the phonological distribution in Turkish will limit our ability to balance the consonant types in the various conditions. However, it should be possible to correlate intrusive vowel quality with consonant place of articulation. In addition to that, we would like to see if surrounding vowels have any effect, as predicted by Clements and Sezer (1982). There is notable variation in realization of liquids between male and female speakers of Turkish, especially with regard to coda /r/, which was pronounced mostly as an approximant [ɹ] by female participants and as a tap or flap [ɾ/ɽ] by male participants, while both groups pronounced it occasionally as a fricative [ɹ̥]. As mentioned earlier, it was noticed that coda /l/ was often pronounced as a fricative [ɬ], which made it easier to delineate vowels before /l/. It would be interesting in future work to investigate the conditions (phonological, phonetic, or sociolinguistic) for this frication of /l/.

We expect that the level of exposure to English affects the extent to which a speaker of Turkish inserts vowels to break onset clusters. Specifically, we expect that speakers who have more direct exposure to English will insert vowels in onset clusters less frequently, and the inserted vowels will be shorter in duration. This is because English allows onset clusters without any vowel insertion, and a lot of the words with onset clusters are loanwords from English where no vowel is inserted. We expect that the quality of the vowel will depend more on region of origin of the speaker, but it is possible that it could be affected by other factors, including knowledge of English. These issues were not able to be addressed in this study due to the low number of participants, but we hope to investigate the source of the attested variation in more depth in the future.

This study found a potential correlation between lexical frequency and rate of insertion in [CC contexts that merits further investigation. In order to investigate frequency effects, there would need to be more stimuli, balanced by category of lexical frequency, e.g. common, uncommon, and nonsense word (personal communication, Barış Kabak, October 8, 2016).

Clements and Sezer (1982) present a phonological model that tries to account for both types of insertion, but the evidence in this paper suggests that insertion in [CC condition is a phonetic phenomenon (intrusion), and is not based in phonology, and in fact the inserted vowel does not even have a phonologically specified articulatory target. Given these conclusions, and the observation that word frequency may be related to

frequency of vowel intrusion in [CC condition, some other model seems necessary. This raises the issue of what such a model would look like. How to model a system that requires epenthesis of a vowel with a phonological target, but that is less strict about enforcing the insertion of an intrusive vowel without a phonological target, while both types of insertion are triggered by the consonantal environment and the latter is affected by word frequency, is left as a goal for future research.

References

- Bellik, J. (2016a). Onset cluster repair in Turkish: an ultrasound study (abstract). Retrieved on February 7, 2016 from http://people.ucsc.edu/~jbellik/pdf/Bellik2016_TurkishUS_abstract.pdf
- Bellik, J. (2016b). Turkish onset-cluster repair: Intrusion or epenthesis? An Ultrasound study. Presented at LASC 2016, University of California, Santa Cruz.
- Bellik, J. (2016c). Turkish onset-repair: an ultrasound study. Graduate qualifying paper, Linguistics Department, University of California, Santa Cruz.
- Boersma, P. & Weenink, D. (2016). Praat: doing phonetics by computer [Computer program]. Version 6.0.14, retrieved February 2016 from <http://www.praat.org/>
- Clements, G. N. & Sezer, E. (1982). Vowel and consonant disharmony in Turkish. In Harry van der Hulst and Norvin Smith (eds.), *The Structure of Phonological Representations*, Part 2, 213-256.
- Göksel, A. & Kerslake, C. (2005). *Turkish: A Comprehensive Grammar*. London: Routledge.
- Hall, N. (2006). Cross-linguistic patterns of vowel intrusion. *Phonology* 23: 387–429. Cambridge University Press.
- Hall, N. (2011). Vowel Epenthesis. In Marc van Oostendorp, Colin J. Ewen, Elizabeth V. Hume, Keren Rice (eds.), *The Blackwell Companion to Phonology*, v. III, Malden, MA and Oxford: Wiley-Blackwell, pp. 1576-1596.
- Hillenbrand, J., Getty L. A., Clark M. J., and Wheeler K. (1995). Acoustic characteristics of American English vowels. In *Journal of the Acoustical Society of America* v. 97 (5), pp. 3099-3111.
- Kabak, B. (2003). *The Perceptual Processing of Second Language Consonant Clusters* (Doctoral dissertation). University of Delaware.
- Levi, S. V. (2001). Glides, laterals, and Turkish Vowel Harmony. In *CLS: The Main Session*, pp. 379-393. The Chicago Linguistics Society.
- Sezer, H. (2016). Processing illicit onset clusters: A case study from Turkish. Poster presented at the Second Workshop on Turkish, Turkic, and the languages of Turkey (Tu+ 2), Bloomington, Indiana.
- Topbaş, S. & Kopkallı-Yavuz, H. (2008). Reviewing sonority for word-final sonorant+obstruent consonant cluster development in Turkish. In *Clinical Linguistics & Phonetics*, v. 22 (10-11), pp. 871-880.

Appendix

Demographic background of participants

	Approx. age	Gender	Time in US	Hometown	Dialect
1 st participant	~30	M	~2 years	Konya	non-Standard
2 nd participant	25-30	F	~2 years	Bursa	Standard-like
3 rd participant	25-30	F	~1 years	Istanbul	Standard
4 th participant	~30	M	~5 years	Adana	Standard

List of all sentences with stimuli

Hasan kabir kelimesini yazdı.
 Hasan Bret kelimesini yazdı.
 Hasan tabir kelimesini yazdı.
 Hasan birim kelimesini yazdı.
 Hasan şekil kelimesini yazdı.
 Hasan klip kelimesini yazdı.
 Hasan tekil kelimesini yazdı.
 Hasan kilim kelimesini yazdı.
 Hasan krem kelimesini yazdı.
 Hasan fikir kelimesini yazdı.
 Hasan bakir kelimesini yazdı.
 Hasan kireç kelimesini yazdı.
 Hasan kadir kelimesini yazdı.
 Hasan drenaj kelimesini yazdı.
 Hasan sedir kelimesini yazdı.
 Hasan dirsek kelimesini yazdı.

List of filler sentences

Hasan okuldan geldi.
 Hasan arabasını sürdü.
 Hasan bu kelimeni yazdı.
 Hasan ödevini yaptı.
 Hasan okula gitti.
 Hasan yemeğini yedi.
 Hasan arkadaşıyla konuştu.

Kazakh Modal Particle *ǰoj*

Nadežda Christopher
SOAS, University of London

Abstract

This paper is the first to be wholly dedicated to describing the Kazakh particle *ǰoj*, its distribution, functions and meaning. It is concluded that, in its pragmatic marker function, the word-form *ǰoj* belongs to the word-class of modal particles; this is a language-specific word class, distinct from other pragmatic markers, such as discourse or text-connective markers. It is also shown that the underlying meaning of this modal particle is givenness and the interaction of this meaning with different contexts and illocutionary forces results in a variety of pragmatic effects, which caused this particle to be considered multi-functional. This paper also shows, however, that not all the functions assigned to *ǰoj* in the literature are in fact performed by the modal particle variant of this word-form; care should be taken to differentiate homophonous word-forms by examining their syntactic and pragmatic contributions to the propositions in which they appear.

Keywords: Kazakh, Turkic, modal particle, pragmatic marker, givenness

1. Introduction

Although modal particles have been receiving an increased amount of academic attention over the past few years, this fascinating word class remains typologically understudied. The data used in the research on modal particles predominantly comes from the Germanic languages (such as German and Dutch), which are remarkably rich with modal particles; little research has been done on the modal particles of the Turkic languages, and, more specifically, on the modal particles of the Kazakh language.

This paper aims to start filling this gap in the typological research on modal particles, and in the larger field of Turkological studies, by providing the first detailed account of the Kazakh modal particle *ǰoj*.¹ Although some information on this particle is available in the literature, a full account of its distribution, functions and underlying meaning has not been proposed; the underlying meaning is the meaning representing the core functional component of *ǰoj*, the interaction of which with different contexts produces a variety of pragmatic effects attributed to this particle.

The paper proceeds as follows: in Section 2, I briefly outline the views on the distinction between modal particles and discourse markers as presented in Degand et al. (2013) and Diewald (2013); Section 3 is dedicated to previous research on *ǰoj*, and demonstrates that there is no consensus on the status of this word-form, its underlying meaning or function. In Section 4 I focus on the distribution of *ǰoj*, provide contextualised examples for the use of this particle, and propose that givenness is its underlying meaning. In Section 5 I present some further observations on *ǰoj*, namely, I posit that it has a heteroseme – the copula *ǰoj*; in the same section I summarise the findings presented in this paper and pose further research questions for the modal particle *ǰoj* and its equivalents in other languages.

The Kazakh language examples used in this paper come from three main sources: my fieldwork in Kazakhstan (September 2015 – May 2016), films and cartoons (namely: *Miñ Bir Tün* (MBT), *Monsters*

¹ The author is grateful to the Wolfson Foundation for the generous PhD Scholarship which made this research possible.

University (MU), and Alice through the Looking Glass (ATLG)), and the National Corpus of the Kazakh Language² (NCKL); the source is given in parenthesis after each example. I also use some examples from other authors – these are explicitly introduced as such with page and the original example numbers provided where possible.

2. Modal vs. Discourse Particle

Modal particles and discourse markers have received a lot of attention in the recent linguistic literature, however, a consensus is yet to be reached on how to tease these ‘fuzzy’ categories apart. In this paper, I follow Degand et al. (2013) and Diewald (2013) in their understanding of the distinction between modal particles (MPs) and discourse markers (DMs).

Both MPs and DMs are multifunctional linguistic expressions, which are thought to indicate some sort of relationship between the proposition and context, or the proposition and the hearer or speaker. Diewald (2013) identifies both these categories as sub-classes of a wider domain of pragmatic markers and posits that while DMs (along with text-connective markers, such as coordinating conjunctions and connective adverbials) are defined by universal functional and formal criteria, MPs are a language-specific word class. This observation is attested by the fact that the class of MPs is not present in Romance languages or in English, although these languages have other means of achieving the same effect (Waltereit 2001). According to Diewald (2013), then, DMs and MPs are “classes on completely different theoretical levels” (2013: 27). Degand et al. (2013: 14) also conclude that DMs and MPs operate on different levels, as they observe that although both classes perform a general indexing function, the former relate the utterance in their scope to the linguistic context, while the latter – to the situational context.

The distributional features identified for DMs include their ability to appear in utterance-initial, -internal, or -final positions, and on their own – that is to say, they are syntactically non-integrated. DMs perform discourse connective functions, as well as various other functions which can be sub-grouped as response signals, segmentation signals, turn-taking signals, hesitation markers, etc. (Diewald 2013).

MPs are syntactically integrated, non-inflecting, and unstressed. They have no constituent or phrasal value, and no referential meaning. In German they are restricted to the middle field (to the right or left of the finite verb), and cannot appear in the first position of a German V-2 sentence, that is, they are restricted to a specific distributional position. Diewald (2013) summarises the main function of MPs as follows:

By using a modal particle the speaker marks the very proposition it is used in as given, as communicatively presupposed, as a particular type of pragmatic presupposition (2013: 33).

Table 1 gives a summary of the main criteria for comparing DMs and MPs as outlined in Degand et al. (2013) and Diewald (2013). I only include those features that have been clearly identified as same or different in these works; some features, such as stress, for example, has only been unequivocally defined for MPs, which are said to be obligatorily unstressed (Diewald 2013: 29).

² The web-address is: <http://web-corpora.net/KazakhCorpus/search/> [Accessed 20 March 2017].

Feature	Discourse Markers	Modal Particles
Constituent Value	No	No
Indexical function	Yes	Yes
Scope	Over non-propositional discourse elements of various sizes	Over propositions or speech-acts
Syntactically integrated	No	Yes
Standalone use	Yes	No
Language-specific	No	Yes

Table 1. Discourse Markers vs. Modal Particles

Table 1 shows that the main criteria whereby a modal particle can be told apart from a discourse marker are their syntactic integration and standalone use. The distribution of *Ǵoj* presented in Section 4 demonstrates that it cannot be a discourse marker and should be viewed and analysed as a modal particle.

In the following Section I show that *Ǵoj* has been referred to as a modal particle, a discourse particle, a modal-expressive particle, an emphatic-limiting particle, and as simply a particle although no grounds are provided for placing *Ǵoj* into any of these categories.

3. Previous Research

Very little has been written about the particle *Ǵoj* (also *ǵoy* and *ǵoy* in the literature) in the English, Russian or Kazakh linguistic literature. This might be due to this particle being perceived (by the native speakers of Kazakh) as an item exclusively appearing in spoken language, and, as such, only carrying some sort of emotive or emphatic meaning without interacting with the grammatical structure of the language.

This attitude is reflected in Straughn (2011), who refers to *Ǵoj* as a sentence-final particle, which expresses emotivity. He claims that *Ǵoj* is completely optional and is never a part of the verbal complex, which is why its presence ‘should not be seen as any sort of formal marking’ (2011: 134). Straughn gives the example in (1) (Straughn’s (190)) to demonstrate that ‘the addition of *ǵoy* merely indicates that the speaker is expressing an emotive attitude toward the content of this utterance’ (2011: 135); in his gloss Straughn labels *Ǵoj* as ‘EXCL’ for ‘exclamative particle’.

- (1) Qıtay-lar-dıń žaǵa žıl-ı eken ǵoy.
 Chinese-PL-GEN new year-3 EVID EXCL
 ‘It’s (apparently) the Chinese New Year!’

Straughn (2011) concludes that the particle *ǵoy* indicates ‘the speaker’s emotive stance’ and is ‘essentially limitless in distribution’, thus it, along with the Uzbek equivalent which Straughn also considers in his thesis, ‘should not be seen as primary verbal markers of emotivity, but instead as discourse particles’ (p. 136). In Section 4 I show that *Ǵoj* is not a limitless in distribution exclamative particle, and that its functions go beyond merely expressing the speaker’s emotive attitude towards the content of his utterance.

Muhamedowa (2016) summarises her analysis of meaning of *Ǵoj* as follows:

It is difficult to find an appropriate translation for this particle in English. *ǵoy* is similar to the English *you know*, as it appeals to shared information between the speaker and the hearer (p.163).

Muhamedowa (2016) also states that this particle has a ‘fixed position in a sentence and must occur after the predicate’ (p.15). As Section 4 shows, Muhamedowa’s description of the placement of *ǰoj* after the predicate is more accurate; the ‘sentence-final’ position proposed by Straughn (2011) does not cover cases where one of the terms appears after the predicate (as an afterthought).

The most detailed description of *ǰoj* I could find in the literature is that presented in Abish’s (2014) doctoral thesis *Modality in Kazakh as spoken in China*. The author refers to *ǰoj* as a modal particle and introduces it as follows:

The particle *I²oy* does not have any lexical meaning. Its basic contribution is to mark some type of epistemic evaluation, an assessment of the propositional content. It can express a commitment to the truth of the proposition, i.e. to its certainty, probability, possibility, etc. The source of the epistemic evaluation can be the opinion of the addresser or some other person. The basic meaning varies according to the communicative functions of different types of usages [...] (2014: 75).

Abish identifies two variants of *ǰoj* – the accented and the unaccented one. The unaccented variant is viewed by the author as an enclitic, which expresses presumption, while the accented variant is used to express repudiation, emphasise shared knowledge, or to form tag-questions (in which case it is pronounced with rising interrogative intonation), as well as to function together with existentials and the conditional mood. I was not able to observe the difference in the accenting of the particle as described by Abish, and some of the examples provided by the author were not attested by the Kazakh speakers in Kazakhstan. This may be due to the differences in how *ǰoj* is used in the Chinese and Kazakh varieties of the Kazakh language. Exploring these differences is outside of the scope of this work, but is an interesting future research question.

Turning to the descriptions of *ǰoj* written in Russian,³ Balakaev et al. (1962) attribute *ǰoj* to both emphatic-limiting and modal-expressive particles (the authors do not provide criteria or features for either of these particle classes); the description of functions of *ǰoj* is only provided under the former heading. Thus, according to Balakaev et al. (1962: 417), the emphatic-limiting particle *ǰoj* is said to express the speaker’s confirmation of his own words. It is used when the speaker wants to remind his interlocutor of an event or action already known to her; additionally, *ǰoj* is used to logically underline or stress a word.

Bol’shoj Kazakhsko-Russkij Slovar’ (henceforth, *The Big Kazakh-Russian Dictionary*) (1998:366) defines *ǰoj* as ‘a particle used to add emphasis or expressivity to the content of an utterance; translated into Russian as *ved’* or *že*, as in (2), or as ‘a particle used to soften a request or a command’, as in (3).

(2) Ajt-ti-m ǰoj!
 say-PAST-1.SG ǰoj
 ‘I did say!/I said, didn’t I?’

(3) Kele ǰoj!
 come.IMP ǰoj
 ‘Come here then!’

Note that the examples in (2) and (3) are provided without contexts, which makes it difficult to establish the conditions in which these utterances would be felicitous; this is especially relevant for the utterance in (2), as it presents the speaker’s reaction to another utterance or an event. I discuss the example in (3) in detail in Section 4.1 and present an alternative analysis of these imperative constructions.

Most Kazakh grammars intended for language learners do not provide a description or explanation of the use of *ǰoj*, and those that do only highlight one aspect of its use. These explanations can be contradictory

³ All translations from Russian are mine unless stated otherwise.

not only between authors, but also between publications by one and the same author. For example, Romanenko (2011: 72) states that *ǰoj* is a particle of “emotional fortification”, while Romanenko (2015: 109) asserts that it is a “clarifying” particle, which “confirms the verity of an utterance”.

Bizakov (2014: 181) refers to *ǰoj* as a “veracity particle”, and Valjaeva (2007-2017) claims that *ǰoj* is a “specifying” particle, which gives an utterance “the meaning of completeness and assertiveness”.

Thus, it is clear that there is no consensus on neither the status, nor the meaning of *ǰoj*. It has been referred to as a particle, a modal particle, a discourse particle, an emphatic-limiting particle, a modal-expressive particle and a veracity particle. Its distribution has been described as limitless and as strictly post-predicative. It has been claimed that *ǰoj* refers to shared information, adds emotivity, confirmativity, assertion, clarity, veracity, specificity, as well epistemic evaluation to an utterance. All these meanings and functions ascribed to *ǰoj* have been illustrated with non-contextualised examples, which do not allow for a full understanding of the contribution the particle makes to a proposition.

In the following Section I take a close look at the distribution of *ǰoj* and provide contextualised example of its use.

4. Distribution of *ǰoj*

The modal particle *ǰoj* is widely used in spoken Kazakh and in informal written styles, and is excluded from formal writing styles. Due to consonant assimilation in Kazakh the particle can be realised as either *qoj* or *ǰoj*; the former is used after voiceless consonants, and the latter is used in all other cases – the initial consonant becomes voiced if preceded by a voiced segment (a vowel or a voiced consonant).

ǰoj possesses all the properties of a modal particle, as outlined in Section 2 – it is unstressed, uninflected, does not form a constituent and cannot form an utterance on its own. Unlike discourse markers, *ǰoj* cannot appear in any position in an utterance. The particle follows predicates, which can be expressed by a lexical or modal verb, an existential or evidential copula, an adjective, or a noun. The examples below show *ǰoj* following a finite verbal predicate in (4), a modal verb in (5), and an existential copula in (6); the example in (4) demonstrates that the predicate might not be in its canonical clause-final position, thus making the particle not obligatorily sentence-final, contrary to Straughn’s (2011) observations. The context – either situational or linguistic – is given in square brackets.

- (4) [One friend utters to another as they hug after not having seen each other for a long time.]
 Fusun, äbden saǰın-dı-q qoj seni.
 Fusun very.much miss-PAST-1.PL ǰoj 2.SG.ACC
 ‘Fusun, we missed you very much indeed.’ (MBT)
- (5) [After finding out that one of the female employees who is not married has a son, one manager says to another:]
 Äkesi bol-u kerek qoj!
 Father-POSS.3.SG be-INF must ǰoj
 ‘There’s got to be a father, hasn’t there/right?’ (MBT)
- (6) [A colleague is late for a morning meeting and the boss says: ‘Do you not have an alarm clock?’ She replies:]
 Zoq, bar ǰoj.
 No EXIST ǰoj
 ‘Yes, of course I do.’ (MBT)

The combination of *ğoj* with the existential copula *bar* is widely used in spoken Kazakh to re-activate a referent about which the speaker wishes to contribute some further information, as shown in (7).

(7) [Talking about upcoming celebrations of Naurız]

Khan Šatır-diñ žanında alañ bar ğoj, onda erteñ koncert
 Khan Shatyr-GEN near square EXST ğoj there tomorrow concert
 bol-a-di.
 be-FUT-3.SG

‘There is this square near Khan Shatyr [shopping centre], right, there will be a concert there tomorrow.’
 (overheard)

This combination of the existential *bar* and particle *ğoj* has been carried over into the Kazakh Russian – the variety of Russian spoken in Kazakhstan – as *est’ že*⁴ (‘there is *že*’) and performs the same function there; this construction is not used in Standard Russian.

Abish (2014) refers to the combination of *bar* and *ğoj* as a complex particle in which *ğoj* is accented. He also claims that this complex particle ‘can follow any constituent of a sentence’ (2014: 83) Firstly, I see no reason to treat *bar ğoj* as one complex particle, as *ğoj* can be omitted from the sentence without causing ungrammaticality, while *bar* cannot be. In the case of omission of *ğoj* only the pragmatic felicity conditions in which the utterance can be successfully used would change – I discuss this in more detail in Section 4.2. Secondly, as mentioned in Section 3, it may be the case that the variety of Kazakh spoken in China differs from Standard Kazakh, however, in Standard Kazakh the particle *ğoj* is not stressed in this construction – the main stress is on the existential *bar*. And lastly, the claim that ‘*bar ğoj*’ can follow any constituent is also not applicable to Standard Kazakh, where the same rules as for a standard existential sentence apply (whether *ğoj* follows or not). Namely, *bar* (as well as its negative counterpart *žoq*) can be preceded by a genitive-possessive nouns sequence (to indicate someone’s ownership of something), or by a locative noun phrase followed by an unmarked noun phrase – as seen in (7).

The example in (8) shows *ğoj* used after an adjectival predicate.

(8) [The opening line of a folk-tale about camels.]

Tüje qazırdiñ özinde ädemi ğoj. [Al erterek-te bu-dan...]
 camel now itself beautiful ğoj but earlier-LOC this-ABL

‘Camels nowadays are beautiful, aren’t they. / As is well known, camels are beautiful at present time. [But a long time ago... (and the tale continues to say that camels used to be more beautiful and how they lost that beauty)].’
 (NCKL)

This example serves as another piece of evidence for classifying *ğoj* as a modal particle rather than a discourse marker as we would not expect a discourse marker to appear in the very first sentence of a tale, where there is no preceding discourse with which to connect.

4.1 *Ğoj* in Imperative and Interrogative Sentences

I start this sub-section with considering *ğoj* in imperative sentences. Muhamedowa (2016: 27) states that ‘the particle *yoy* attached to the –A converb softens an imperative’, and provides the following example (the author’s formatting is preserved, the footnote is mine):

⁴ Diana Forker reports the same for the variety of Russian spoken in Dagestan. This can be explained by the similar use of the Sanzhi Dargwa particle =*q’al* – which is comparable to *ğoj* and *že* in its functions (Forker 2017).

- (13) Erteñ sağat bes-te žeñgejdi ert-ip,
 tomorrow hour five-LOC sister-in-law-ACC bring.with-PTCPL
 biz-diñ üj-ge kel-e qoj-iñiz.
 1PL-POSS house-DAT come-CONV AUX.IMP-3.PL

‘At 5 o’clock tomorrow, please come to our house together with the sister-in-law.’ (NCKL)

I hope to have convincingly shown that what has been reported as the modal particle *ğoj* being used to ‘soften’ commands is in fact the imperative form of the verb *qoü* used as an auxiliary verb together with the converbial form of the lexical verb. This, however, is not to rule out the possibility of a modal particle like *ğoj* appearing in an imperative sentence, as this is possible for the Russian *že* (*Idi!* – ‘Come.IMP.2SG’ and *Idi že!* – ‘Come.IMP.2SG then / Please come!’). It may well be the case that the analyses of *ğoj* in imperatives presented in the the *Big Kazakh-Russian Dictionary* (1998) and in Muhamedowa (2016) were influenced by the Russian language due to the past socio-linguistic situation in which Russian was the dominant language of the Kazakh Soviet Socialist Republic.

Let us now turn to the modal particle *ğoj* in interrogative sentences. As MPs indicate that the proposition in which they appear is pragmatically presupposed, we cannot expect them to appear in the true interrogative sentences, i.e. questions requesting new, non-presupposed information. This is indeed the case for *ğoj*, as shown in (14), (15) and (16).

- (14) *Sen barasıñ ba ğoj?
 2.SG go-FUT-2.SG QUEST ğoj
 Intended: ‘Are you going?’ (elicited)
- (15) *Sen barasıñ ğoj ma?
 2SG go-FUT-2SG ğoj QUEST
 Intended: ‘Are you going?’ (elicited)
- (16) *Ol qaşan kel-di ğoj?
 3SG when come-PAST.3.SG ğoj
 Intended: ‘When did he come?’ (elicited)

However, it is important to note, that although this particle cannot occur in true interrogative clauses (those containing *wh*-questions or question particles), it is widely used in utterances with interrogative prosody (rising intonation), and pragmatic force of questions, as in (17).

- (17) Sen bügin universitet-ke bar-a-siñ ğoj?
 2SG today university-DAT go-FUT/PRES-2.SG ğoj
 ‘You are going to university today, aren’t you?’
 ‘You are going to university today, right?’ (elicited)

The possible responses to (17) would sound the same as the responses to a question formed with the question particle *MA* – ‘Yes, I am going’ or ‘No, I am not going’. However, if the utterance in (17) is pronounced with a falling intonation on the particle, it can be felicitously used as a response to a question as shown in (18), while a true interrogative sentence could not be felicitously used here, which indicates that the pragmatically interrogative utterances with *ğoj* are not true questions.

- (18) Speaker 1: Why did you say you can’t leave the house today?
 Speaker 2: Sen bügin universitetke barasıñ ğoj, [and there is no one else to look after the children].

Speaker 2': #Sen bugin universitet-ke barasın ba?
 2.SG today university-DAT go-FUT/PRES-2.SG QUEST
 'Are you going to university today?' (elicited)

The utterances with *Ǵoj* in (17) and (18) differ only in prosody, more specifically, in the intonation on the final element – the particle itself. Interestingly, the same is observed for the English tag-questions (cf. Rando 1980, Huddleston and Pullum 2002, Reese and Asher 2006), which are the closest translation equivalents for utterances like those in (17) and (18). Abish (2014) refers to this use of the modal particle as 'non-modal', although no detailed explanation is provided for this description.

To sum up, the modal particle *Ǵoj* is unstressed, uninflected and cannot form an utterance on its own. It follows verbal or non-verbal predicates. It does not appear in imperative or interrogative sentences, although it can create pragmatically interrogative utterances, the nearest equivalents of which in the English language are tag-questions.

4.2 Meaning of *Ǵoj*

Ǵoj is a multi-functional particle which is used in a variety of contexts without contributing to the truth-conditional content of utterances in which it appears; the examples below show *Ǵoj* being used in different contexts.

- (19) [Parents check up on their sleeping child. The Mother says to the Father:]
 Kör-di-n be? Ujktap žatir dep ajttim Ǵoj.
 See-PAST-2SG QUEST sleep-CONV AUX.3SG COMP say-PAST-3SG Ǵoj
 'Did you see? I did tell you he was sleeping.' (MU)
- (20) [On a school trip the teacher finished counting the children as they get off the bus. She counts 19, but there should be 20 students.]
 Sonda, bireu žoq qoj.
 then someone NEG.EXST Ǵoj
 'Someone is missing, aren't they?' (MU)
- (21) Speaker 1: Qajrat keše düken-ge bar-ğan žoq.
 Kairat yesterday shop-DAT go-PAST.PTCPL NEG
 'Kairat did not go to the shop yesterday.'
 Speaker 2: Ol bar-dı⁶ Ǵoj!
 3.SG go-PAST.3 Ǵoj
 'He did go though! / Of course he went! / He did too!' (elicited)
- (22) [After asking a question and not receiving an answer from his wife, the man utters with annoyance:]
 Men sura-p tur-dı-m Ǵoj!
 1SG question-CONV AUX-PAST-1.SG Ǵoj
 'I asked (you) a question, didn't I?' (MBT)

⁶ The syllable in bold is stressed. In the same clause without *Ǵoj* the stress would fall on the second syllable – bar-dı. This paper does not deal with this stress shift, but it shows that *Ǵoj* is not a frivolously added item, but interacts with the rest of the proposition on all levels, including prosody.

- (23) [After having been told that Alice was on her way and having waited for her to arrive, the man exclaims:]

Ol kel-me-j-di ġoj?!

3.SG come-NEG-FUT/PRES-SG ġoj

‘She is not coming, is she?’

(ATLG)

In all of the examples from (19) to (23) the use of the modal particle *ġoj* indicates that the information contained by the proposition in its scope is either given, or communicatively or pragmatically presupposed, and the utterances containing *ġoj* either point out this givenness or reactivate the information in the Common Ground shared by the interlocutors. However, some additional, narrower meanings, which can be attributed to the presence of *ġoj*, can be observed in the examples (19) - (23). Thus, in the example in (20) there is an effect of exclamation and surprise at the speaker’s own realisation that someone was missing which is re-enforced by the addition of *ġoj*. The modal particle scopes over the proposition *bireu žoq* and marks it as ‘given’, thus re-confirming it. In a way, we can describe this as a double verbalisation of the situation: first, the speaker describes the state of affairs in her proposition *bireu žoq*; then, she gives the proposition an epistemic marking with the modal particle *ġoj* to show that the state of affairs described by the proposition is obvious or given in the situational context.

In the second utterance in (21) the additional effect of the use of *ġoj* is the expression of reproach; not only is *ġoj* used to point to the fact that the information provided by Speaker 2 was accessible to Speaker 1, but also to show disappointment or disapproval that this information had not been used. This adversative use of *ġoj* with verum focus is fairly frequent in corrective utterances. Omission of *ġoj* would result in the same utterance semantically, but the pragmatic effect of givenness and reproach would not be conveyed.

In (22) we see *ġoj* being used to create the effect of a rhetorical question, which is uttered to show the speaker’s annoyance. Interestingly, in this example the speaker does not refer to the shared knowledge per se, but rather to his own action (‘asking a question’) which happened only a few moments prior; the addition of *ġoj*, which carries the meaning of givenness, highlights the fact that the hearer was present when the original question was uttered but chose not to reply to it. The speaker could have simply restated his question, but by uttering (22) he shows his annoyance at how the exchange has unfolded so far.

In (23) we see the tag-question effect created by the use of *ġoj* in a similar way as in (17). The utterance in (23) is uttered by someone who had been told that Alice was on her way, however, having waited for quite some time it became obvious that she was not coming at all. By producing (23) the speaker states the obvious – the fact that she was not coming –, and indicates the givenness of this proposition by using *ġoj*; the tag question effect is created by the rising intonation on *ġoj*. We can suppose that the givenness or re-activation meaning of *ġoj* combined with the interrogative intonation create the effect along the lines of: *p* is known/evident/given, is it still the case that *p*? This is also applicable to the example in (17).

Let us briefly return to the example in (7), repeated below in (24) for convenience, and examine the contribution of *ġoj* there.

- (24) [Talking about upcoming celebrations of Nauriz]

Khan Šatir-diñ žaninda alaň bar ġoj, onda erteň koncert

Khan Shatyr-GEN near square EXST ġoj there tomorrow concert

bol-a-di.

be-FUT-3.SG

‘There is this square near Khan Shatyr [shopping centre], right, there will be a concert there tomorrow.’
(overheard)

As before, *ġoj* reactivates and points to givenness of the proposition it scopes over, in this case, that there is a square next to the Khan Shatyr shopping centre. This construction serves the goal of introducing a topic by re-activating a referent which is presumed to already exist in the hearer's mental representation of the world. This utterance would not be pragmatically felicitous if the speaker knew for sure that the hearer is not aware of the square next to the Khan Shatyr shopping centre (perhaps someone who has not lived in Astana for a long time); in this case *ġoj* would be omitted from the utterance, which would then be produced and parsed as two sentences – the first would inform the hearer that there is a square next to the Khan Shatyr shopping centre, and the second sentence would tell her about the concert.

In summary, the main claim of this section is that the underlying and unifying meaning of the modal particle *ġoj* is givenness in its broadest sense possible. The information carried by the proposition might be 'given' as a result of previous interaction between the interlocutors, due to assumed previous knowledge or experience, or due to non-linguistic context. This underlying meaning of givenness results in different pragmatic effects when used in different contexts and with different illocutionary forces, thus making the modal particle *ġoj* a truly multi-functional particle.

5. Further Observations, Questions and Concluding Remarks

In this section I present some additional observations on the modal particle *ġoj* and its heteroseme – copula *ġoj*. I then summarise the outcomes of this paper and pose questions for future research on the modal particle *ġoj* and its equivalents in other languages.

5.1 Heteroseme of *ġoj*

Diewald (2013) identifies the existence of heterosemes in other word classes as one of the core features of the German MPs. In this section I show that the modal particle *ġoj* has a heteroseme which seems to perform the function of a copula.

The copula *ġoj* appears in cleft constructions, where it follows the contrastively focused non-verbal element, as shown in (25) below. It seems to perform the copula-auxiliary function here, as sentences with participial verb-forms would not make a grammatical main clause on their own. Essentially, *ġoj* turns a subordinate clause into a main clause⁷ and omitting *ġoj* would result in an unacceptable independent clause.

- (25) Speaker 1: Keše Bolat düken-ge bar-dï.
 yesterday Bolat shop-DAT go-PAST.3
 'Bolat went to the shop yesterday.'
- Speaker 2: Qajrat *(qoj) bar-ġan/ *bar-dï!
 Kairat COP go-PAST.PTCPL/ *go-PAST.3
 'Of course it was Kairat who went!'
- Speaker 2': Zoq, ol kitaphana-ga *(ġoj) bar-ġan/ *bar-dï!
 No 3SG library-DAT COP go-PAST.PTCPL/ *go-PAST.3
 'No, it was the library that he went to.' (elicited)

⁷ Forker (2017) notes the same for the Dargwa particle =*q'al* (also mentioned in Footnote 2). She refers to this use of =*q'al* as predicative particle which is a word class identified in the Dargwa languages, and which serves functions that are usually performed by copula-like auxiliaries. Such a word class has not been identified for Kazakh and examining this language for this word class goes far beyond the scope of this paper, however, it might be an interesting question for further research.

The same is observed in the present tense, as can be seen in (26) below.

(26) [Two friends discussing two brothers called Qajrat and Bolat. The first speaker claims that Qajrat is the one who goes to the library the most. The second speaker objects:]

Qoj-ši, Bolat *(qoj) eñ köp bar-atın / *bar-a-di
 Stop-POL Bolat COP most a.lot go-PRES.PTCPL go-PRES-3.SG
 ‘Oh stop! Of course it is Bolat who goes to the library the most!’ (elicited)

Although *ğoj* still carries the pragmatic meaning of givenness (as it is described in the previous section) in these examples, its distribution and syntactic functionality differs from that of a modal particle in that: a) it follows and scopes over a constituent, rather than the whole proposition, and b) it directly participates in the syntactic structure of the clause and turns a dependent clause into an independent one.

Another example of copulaic use of *ğoj* is shown in (27) below, where it functions as a link between a speaker’s utterance and his own epistemic evaluation of it.

(27) [Upon seeing a book in a shop, a person might utter to a friend.]

Osı Marat-tiñ kitabı *(ğoj) de-j-min.
 This Marat-GEN book-2.SG.POSS COP say/think-PRES-1SG
 ‘It seems to me that this is Marat’s book.’ (elicited)

As previously, the omission of *ğoj* results in ungrammaticality. There appears to be a degree of epistemic modality associated with the combination of *ğoj* and the verb *deu* (‘to say’; ‘to think’) when used as shown above; the speaker indicates that he is not entirely certain of the proposition expressed by the subordinate clause.

Further research is necessary to fully understand the behaviour and functions of the copula *ğoj*, however, the examples given above are sufficient to identify this use of *ğoj* as distinct from the heterosemous modal particle *ğoj*.

5.2 Concluding Remarks and Further Questions

Three main conclusions have been drawn in this paper. The first conclusion concerns the nature of the pragmatic marker *ğoj*. Based on Degand et al. (2013) and Diewald (2013) distinctions of discourse markers and modal particles, it has been conclusively shown that Kazakh *ğoj* is a modal particle. Unlike a discourse marker, the modal particle *ğoj* cannot form an utterance on its own, and is confined to the position immediately after the predicate (even when the predicate is not sentence-final).

The second conclusion is to do with the underlying meaning of *ğoj*. I propose that the underlying meaning of this modal particle is givenness (in its broadest sense), and that the possibility of combining this meaning with different illocutionary forces, moods, as well as other linguistic and non-linguistic contexts explains the multi-functionality of this modal particle.

The third conclusion is that not everything that looks and sounds like the modal particle *ğoj* is the modal particle *ğoj*. It was shown that what was claimed to be the modal particle *ğoj* in imperative sentences is, in fact, the imperative form of the auxiliary verb *koü*. It was also shown that *ğoj* has a heteroseme which performs the function of a copula, while still carrying the pragmatic meaning of givenness.

One of the main further questions which arises from the observations presented in this paper is the possibility of a formal syntactic representation of the contribution the modal particle *ğoj* makes to a proposition in which it appears. Due to the pragmatic nature of this contribution, its formal representation seems to only be possible in syntactic frameworks which already possess or allow for the development of tools for linking

propositions with (linguistic or non-linguistic) context. Dynamic Syntax is one such theoretical framework and further research on modal particles and other pragmatic markers, as well as on their interaction with different contexts in which they appear seems particularly promising.

Another direction in which further research on ğoj and other modal particles needs to develop is of typological and comparative nature. As mentioned in Section 3, ğoj can be translated into Russian as either *že* or *ved'*, which seems to indicate that ğoj is more multi-functional than *že* or *ved'*. The same can be observed for ğoj and its German counterparts, where a few different pragmatic markers could be used in place of ğoj in different contexts. Further insight into the nature of modal particles could be gained by investigating the interpretation different modal particles might receive in different languages without this word class. This cross-linguistic comparative research could help deepen our understanding of different categories of pragmatic markers in general, and modal particles in particular.

List of Abbreviations

SG	singular	POSS	possessive
PL	plural	GEN	Genitive case
PAST	past	POL	polite
CONV	converb	EXST	existential
PTCPL	participle	FUT	future
DAT	Dative case	QUEST	question particle
ACC	Accusative case	COMP	complementizer
COP	copula	NEG	negation
PRES	present	AUX	auxiliary
IMP	imperative	INF	infinitive
LOC	Locative case	1	first person
ABL	Ablative case	2	second person
EVID	evidential	3	third person

References

- Abish, A. (2014). *Modality in Kazakh as spoken in China*. Ph.D. Uppsala University.
- Balakaev, M.B., Baskakov, N.A., & Kenesbaev, S. K. (1962). *Sovremennij Kazakhskij Jazik*. Alam-Ata: Izdatel'stvo Akademii Nauk Kazakhskoi SSR.
- Bizakov, S. (2014). *Kazakhskij Jazik: Prosto o Složnom (The Kazakh Language: Simply about the Difficult)*. Almaty: Eltanım.
- Bol'shoj Kazakhsko-Russkij Slovar'* (1998). Almaty: Slovar'.
- Degand, L., Cornillie, B., & Pietrandrea, P. (2013). Modal particles and discourse markers: Two sides of the same coin? In: Degand, L., Cornillie, B., & Pietrandrea, P. (eds.), *Discourse markers and modal particles: Categorization and description*, pp. 1-18. Amsterdam: John Benjamins Publishing Company.
- Diewald, G. (2013). "Same same but different" – Modal particles, discourse markers and the art (and purpose) of categorization. In: Degand, L., Cornillie, B., & Pietrandrea, P. (eds.), *Discourse markers and modal particles: Categorization and description*, pp. 19-45. Amsterdam: John Benjamins Publishing Company.
- Forker, D. (2017). *What Relevance Theory can teach us about modal particles: A lesson from Sanzhi Dargwa*. (Submitted for publication in March 2017).
- Haspelmath, M. (1995). The converb as a cross-linguistically valid category. In: Haspelmath M. & König, E. (eds.), *Converbs in Cross-linguistic Perspective*, pp. 1-55. Berlin and New York: Mouton de Gruyter.
- Huddleston, R. & Pullum, G. K. (2002). *The Cambridge Grammar of the English Language*. Cambridge: Cambridge University Press.
- Muhamedowa, R. (2016). *Kazakh: A Comprehensive Grammar*. Oxon: Routledge.

- Rando, E. (1980). Intonation in discourse, In: Waugh, L.R., van Schooneveld, C.H. & Bolinger, D. (eds), *The Melody of Language*, pp. 243–278. Baltimore: University Park Press.
- Reese, B. & Asher, N. (2006). Prosody and the Interpretation of Tag Questions. In: Puig-Waldmüller, E. (ed.), *Proceedings of Sinn und Bedeutung 11*, pp. 448-462. Barcelona: Universitat Pompeu Fabra, pp.448-462.
- Romanenko E.I. (2011). *Kazakhskaja Grammaticka dlja Russkojazichnih (Kazakh Grammar for Russian Speakers)*. Almaty: A-level.
- Romanenko E.I. (2015). *Sovremennij Kazakhskij (Modern Kazakh)*. Almaty: Kazbooka.
- Straughn, C. (2011). *Evidentiality in Uzbek and Kazakh*. Ph.D. The University of Chicago.
- Valjajeva, T. (2007-2017). *Kazakhskij Jazik*. [online] Available at: <http://kaz-tili.kz> [Accessed 15 March 2017].
- Waltereit, Richard. 2001. Modal Particles and their Functional Equivalents: A Speech-Act-Theoretic Approach. *Journal of Pragmatics* 33 (9), 1391–1417.
- Yu Cho, Y-M. (1990). Syntax and Phrasing in Korean. In: Inkelas, S. & Zec, D. (eds.), *The Phonology-Syntax Connection*, pp. 47-62. Chicago and London: The University of Chicago Press.

Is the *-ar/-mas* Participle a Participle in Uzbek?

Shinsuke Hidaka
Tokyo University of Foreign Studies

Abstract

In this paper, I discuss whether the *-ar/-mas* future participle-formed suffix (hereafter the *-ar/-mas* participle) is indeed a participle in Uzbek. Previous works have considered the *-ar/-mas* participle as a participle. However, it has low productivity in comparison to other participles. In addition, previous works suggest that a clause formed by the *-ar/-mas* participle cannot have the force of a verb (for example, such clauses cannot have arguments and temporal adverbs). I, therefore, analyze the syntactic structure of the *-ar/-mas* participle in terms of two basic functions possessed by other participles: (1) clause nominalization, and (2) attributive clause. Based on the above, I conclude that the *-ar/-mas* participle is not a participle in Uzbek, but rather one of the derivational suffixes.

Keywords: Uzbek, participle, derivation

1. Introduction

In this paper, I discuss whether or not the *-ar/-mas* future participle-formed suffix (hereafter the *-ar/-mas* participle) is indeed a participle in Uzbek. Previous works (Asqarova and Jumaniyozov 1953, Kononov 1960: 239, Abdurahmonov et al. 1975: 514, Bodrogligeti 2003: 632) have considered the *-ar/-mas* participle as a participle. However, from the result of my examination, I conclude that the *-ar/-mas* participle is not a participle in Uzbek, but rather one of the derivational suffixes.

2. Previous Works

In previous works, it has been claimed that a participle has two functions. One of them is the attributive function.

- (1) Attributive function

[<i>falokat</i>	<i>yuz</i>	<u><i>ber-gan</i></u>	<i>joy</i>
disaster	surface	give-PTCP.PAST	place

‘the location where the accident took place’ (Bodrogligeti 2003: 617)

The other function is clause nominalization.

- (2) Clause nominalization

[<i>Siz-ni</i>	<i>hechkim</i>	<u><i>aybla-yotgan-i</i></u>	<i>yo‘q.</i>
2PL-ACC	anybody	blame-PTCP.PROG-3SG.POSS	no

‘No one is blaming you.’ (Bodrogligeti 2003: 624)

(lit. ‘None blame you.’)

Table 1 lists the forms of Uzbek’s participles.

	Tense, Aspect etc.	Note
<i>-gan/ -kan/ -qan</i>	Past tense, perfect (Bodrogligeti 2003: 616, Kononov 1960: 238)	None
<i>-(a)yotgan</i>	Present tense (Kononov 1960: 238) An actual present action in progress (Bodrogligeti 2003: 622-623)	< <i>yot-gan</i> [lie-PTCP.PAST] (Kononov 1960: 238)
<i>{-a/ -y}digān</i>	Future-present tense (Kononov 1960: 238, Bodrogligeti 2003: 620)	< <i>tur-gan</i> [stand-PTCP.PAST] (Kononov 1960: 238)
<i>-(a)r</i>	Future tense (Kononov 1960: 239, Abdurahmonov et al. 1975: 514) The subject carries out or will carry out a habitual or eventual action. (Bodrogligeti 2003: 632)	• relatively rare. • negative form: <i>-mas</i> (Kononov 1960: 239)

Table 1. Uzbek's participles

Following this section, I term *-gan*, *-yotgan*, *-digān* as “*-gan* group participles” for the comparison with the *-ar/-mas* participle.

3. Posing the Problem

Previous works (Asqarova and Jumaniyozov 1953, Kononov 1960: 239, Abdurahmonov et al. 1975: 514, Bodrogligeti 2003: 632) have regarded the *-ar/-mas* participle as a participle. However, it has low productivity in comparison to other participles. Kononov (1960: 239) states that the use of this participle is uncommon. Abdurahmonov et al. (1975: 514) state that *kel-adigān* [come-PTCP.NPST] in (3) a. cannot be replaced with *kel-ar* [come-PTCP.FUT], as in (3) b.

- (3) a. *ertaga* *kel-adigān* *odam*
tomorrow come-PTCP.NPST person
- b. **ertaga* *kel-ar* *odam*
tomorrow come-PTCP.FUT person
- ‘The person who will come tomorrow.’

Judging from this description, I believe the *-(a)r/-mas* participle does not take an adverb.

In the case of the attributive function, Abdurahmonov et al. (1975: 514) state that the *-(a)r/-mas* participle is closer to an adjective than to a verb. I believe that this description is supported by the fact that the *-(a)r/-mas* participle does not retain arguments, as in (4) and (5).

- (4) *oq-ar* *suv*
flow-PTCP.FUT water
‘the flowing water’
- (5) *so‘n-mas* *hayot*
vanish-PTCP.FUT.NEG life
‘the non-vanishing life’

(Abdurahmonov et al. 1975: 514)

However, *-gan* group participles retain the direct object and the adverb. In (2), *-yotgan* retains the direct object *Siz-ni* [2PL-ACC]. In (3)a, *-adigan* retains the adverb *ertaga* ‘tomorrow’. If *-gan* group participles are regarded as prototypical, is *-ar/-mas* a “participle”?

4. Examination

I will analyze the syntactic structure of the *-ar/-mas* participle in terms of two basic functions possessed by the other participles, namely, 1) clause nominalization and 2) attributive clause. Hence, I reaffirm Bodrogligeti’s description. Bodrogligeti (2003: 245-246) presents headless relative examples by *-ar/-mas*.

4.1 Clause nominalization

Firstly, I carry out the examination of the corpus¹. This corpus contains 140 examples of *-gan* group participles. However, there is **no** example in which *-ar/-mas* takes an argument and is nominalized, as in point (2). In addition, I asked a native speaker whether *-yotgan* in (2) can be exchanged for *-ar/-mas*. He answered that this is not permitted (see point (6)). That is, *-ar/-mas* cannot nominalize a clause.

- (6) **[Siz-ni hechkim aybla-ar-i]* yo‘q.
 2PL-ACC anybody blame-PTCP.FUT-3SG.POSS no
 [No one will blame you.] (cf. (2))

4.2 Attributive clause

In previous works (especially, Abdurahmonov et al. 1975: 514), the attributive clause by *-ar/-mas* basically cannot retain arguments and adverbs. I regard points (4) and (5) as examples that do not retain an argument, and point (3) as an example that does not retain an adverb.

However, there are a few examples that do retain adverbs and arguments. Firstly, I present some *-ar/-mas* examples that do retain arguments like (7). This example has been quoted from a *folktale*.

- (7) *Hurmat qil-ib sen-ga ayt-ar so‘z-im bor.*
 respect do-CVB 2SG-DAT say-PTCP.FUT word-1SG.POSS existence
 ‘I say this word out of respect for you.’ (Asqarova va Jumaniyozov 1953: 13)

Next, I show an example from my corpus data in which the adverb appears to be retained. This example is part of a *poem*.

- (8) *Abad so‘l-mas muhabbat-ga fido bo‘l-ib o‘t-yap-man,*
 forever droop-PTCP.FUT.NEG love-DAT sacrifice be-CVB pass-PROG-1SG
 ‘I have been sacrificed for love, which never dies.’

Examples mentioned (7) and (8) may be instances of rhetorically stylistic usage owing to the fact that these are used in a folktale and a poem, respectively.

In addition, I show another example from my corpus data in which the adverb appears to be retained. *tez+yur-ar* in (9) has been spelled as one word like *tezyurar* in the original data.

¹ My corpus data is composed of two kinds of text: The first one is a set of 42 articles chosen arbitrarily from the news website *Ozodlik radios* (<http://www.ozodlik.org>; 2014/7~8, 2015/7~8); the other is the novel *Besh qiz va bir yigit* (Five girls and one young man). Every morpheme in these texts has been analyzed.

- (9) *Toshkent-dan Qarshi-ga tez+yur-ar poezd qatna-y boshla-di-ø*
 NAME-ABL NAME-DAT fast+move-PTCP.FUT train go.and.come-CVB start-PAST-3SG
 ‘The express train begins to go and come from Tashkent to Qarshi.’

4.3 Bodrogligeti’s description

Bodrogligeti (2003: 245-246) says that “the second element is the 3rd person singular of the aorist” as (10), (11).

- (10) *kun+chiq-ar*
 sun+go.out-PTCP.FUT
 ‘east’
- (11) *qo’y+boq-ar*
 heep+raise-PTCP.FUT
 ‘shepherd’

He regards *-ar/-mas* in these cases as finite verbs. However, is this true? It may not be, think so, because these examples have a headless relative clause construction. These literal translations are ‘(the place where) the sun rises’ in (10) and ‘(the person who) keeps sheep’ in (11). *-gan* group participles have such a headless relative clause construction as the example in point (12).

- (12) *o’l-gan-lar*
 die-PTCP.PAST-PL
 ‘dead people’ (lit. died)

However, *-ar/-mas* is different from *-gan* group participles. *-ar/-mas* involving an argument is regarded as **one word phonologically and grammatically**. In the following, I show three reasons for this:

1. **qo’yboqar** ‘shepherd’ in (11) takes one stress. Generally, this falls on the final syllable of a word.
2. **qo’y** ‘sheep’ in (11) cannot take any nominal suffix.

- (13) a. **qo’y-ni-boq-ar*
 sheep-ACC-raise-PTCP.FUT
- b. **qo’y-lar-boq-ar*
 sheep-PL-raise-PTCP.FUT
- c. **qo’y-im-boq-ar*
 sheep-1SG.POSS-raise-PTCP.FUT

3. **qo’y** ‘sheep’ in (11) cannot be modified by an adjective. If you want to express ‘the person who raises **fat sheep**,’ you must use *-adigan* as in (14)b.

- (14) a. *semiz qo’y-boq-ar*
 fat sheep-raise-ar
 ‘a fat shepherd’
 (not ‘the person who raises **fat sheep**’)

- b. *semiz qo‘y boq-adigan odam*
 fat sheep raise-PTCP.FUT person
 ‘the person who raises *fat sheep*’

Headless relative clauses using *-ar/-mas* are less productive, as shown in (10) and (11). The arguments and verb taken by this structure are very limited. The Uzbek dictionary (Begmatov va boshq. 2008) registers *kunchiqar* as a head.

5. Conclusion

I present the result of the Examination in Table 2.

	Clause nominalization	Attributive (clause)	
		Retaining argument and adverb	
<i>-gan</i> group	+	+	+
<i>-(a)r/-mas</i>	-	-	+

Table 2: The difference between *-gan* group participles and *-ar/-mas*

In previous works, *-ar/-mas* has been considered a participle. However, these works also say that *-ar/-mas* has low productivity. In my examination, *-ar/-mas* cannot be nominalized and does not, basically, take arguments and adverbs. There are examples in which they have been considered as one word phonologically and grammatically. Finally, I conclude that *-ar/-mas* is not a participle, but a lexical derivational suffix.

List of Abbreviations

-	affix boundary	FUT	future
+	compound boundary	NAME	proper noun
1	first person	NEG	negative
2	second person	NPST	non-past
3	third person	PAST	past
ABL	ablative	POSS	possessive
ACC	accusative	PROG	progressive
CVB	converb	PTCP	participle
DAT	dative	SG	singular

References

Abdurahmonov, G‘. A. va Sh. Sh. Shoabdurahmonov, A. P. Hojiyev (1975). *O‘zbek tili grammatikasi I-tom Morfologiya*. [The reference grammar of Uzbek vol.I Morphology] Toshkent: O‘zbekiston SSR «Fan» nashriyoti.

- Asqarova, M. va R. Jumaniyozov (1953). *O‘zbek tilida ravishdosh va sifatdoshlar Hozirgi zamon o‘zbek tili kursidan materiallar*. [Converbs and participles in Uzbek from the Modern Uzbek class materials] Toshkent: O‘zSSSR fanlar akademiyasi nashriyoti.
- Begmatov, E., A. Madavaliyev, N. Mahkamov, T. Mirayev, N. To‘xliyev, E. Umarov, D. Xutoyberganova, A. Hojev. (2008). *O‘zbek tilining izohli lug‘ati: 80000 dan ortiq so‘z va so‘z birikmasi: 5 jild*. Toshkent: “O‘zbekiston milliy ensiklopediyasi” Davlat ilmiy nashriyoti.
- Bodrogligeti, Andràs J. E. (2003). *An academic grammar of Modern Literary Uzbek*. München: Lincom Europa.
- Kononov, A. N. (1960). *Grammatika sovremennogo Uzbeks-kogo literaturnogo iazyka*. [The reference grammar of Modern literary Uzbek] Moskva, Leningrad: Izdatel’s-tova akademii nauk SSSR.
- Sjoberg, F. Andrée. (1963). *Uzbek Structural Grammar*. Uralic and Altaic Series, Vol.18. Bloomington: Indiana University.

Verbal Aspect Problems and the Aspect of the Kazakh Language

Fazira Kakzhanova
Buketov Karaganda State University

Abstract

The category of aspect is one of the most complicated problems of linguistics. The Kazakh language is considered to be temporal not aspectual, that is why the category of aspect is absent from the chapter on verb morphology in Kazakh grammar textbooks, in spite of having a set of aspects and sub-aspects.

The paper aims at stating the thesis that the Kazakh language has the aspect category. For argumentation of this thesis, three objective reasons, confirming the existence of the aspect category: 1. verb seme generating six verb categories, 2. the plane of aspect content, and 3. the plane of expression are presented.

There are three subjective reasons facilitating to deny the existence of the aspect in Kazakh language: 1. the problems of the plane of aspect expressions: 'absence' of *synthetical* plane of aspect expression and difficulties of acceptance of *analytical* planes of aspect expressions in Kazakh language, which has a high ratio of affixation per a word; 2. 'absence' of 'Aktionsart' in the Kazakh language on the basis of understanding of Slavic aspect theory, because of absence of aspect formation prefixes; 3. the problem of the aspect definition.

And about thirty 'tenses' of the Kazakh language instead of three.

Keywords: aspect, Aktionsart, single continuum of expression, plan of content, sub-aspects

1. Introduction

The category of verb aspect is one of the six categories of the verb. It is one of the most complicated issues of linguistics, the difficulty is that some languages are considered to be aspectual and some languages are temporal, including the Kazakh language. It means, that the Kazakh language has no the verb aspect category in its matrix. The article is about existence of the aspect category in the Kazakh (Turkic languages), in spite of denying of the existence of the aspect in these languages.

The main thesis of this article is that *the Kazakh language is aspectual*.

There are three *objective* reasons confirming about existence of the aspect category in all languages, including the Kazakh language also. To have the aspect category, languages should have:

1. a verb, generating six verbal category meanings;
2. the plane of aspect content (PC);
3. the plane of aspect expression (PE) [1].

The first item is to have the aspect category, languages should have verbs, which can generate six verb category meanings. Verbs are a complex part of speech, which have six (at least known) independent category meanings: aspect, tense, mood, voice, number, and person. The main peculiarities of these verb categories are

that each of them has its own independent discrete meaning (DM) and these category meanings don't overlap or never interfere in meanings of each other, and never participate in the creation of other verb category meanings. Their discrete meanings are:

- aspect expresses modification of an action on the basis of inner stages of action development;
- tense expresses location of an action in the time line;
- voice expresses the agent of the action: subject or object;
- mood expresses the relation of the action to reality;
- number expresses the number of subjects by whom the action is/are realized;
- person expresses the indexical relationship between the agent of the action and the speaker . [2].

A verb seme has substantial properties of generating these verb category meanings. All these verb category meanings are put in the matrix of verbs. Actions as the planes of content of verbs are identified by these enumerated properties and languages should reflect and express these listed properties of action. Let us see how the Kazakh language organizes the plan of expression of these verb category meanings.

2. The plane of expression of these verb category meanings in the Kazakh language

The second objective reason of having the aspect category in languages is to have the plane of aspect expressions. The Kazakh language has two types of planes of expression of the aspect category. They are:

- analytical,
- synthetical.

The Kazakh language is considered to belong to agglutinative languages, having high ratio of affixation per a word as agglutinative language, the Kazakh language expresses the verb category meanings, including the aspect category externally by analytical way, which was difficult to accept by linguists at once, especially the grammaticalization of auxiliary verbs for expression of the verb category meanings.

All these verbal categories function within a single continuum of expression, because they have no their own separate plane of expression. But each of them has some elements in analytical or synthetical continuums of expression expressed by predicates. For example, the analytical predicate *okyp boladu* in the Kazakh language has two components: the gerund *okyp* derived from the main verb *oky* and the auxiliary verb *bol* and plus affixations ‘- a, - du’. Six verb category meanings are distributed in the predicate in such a way:

<i>okyp</i>	<i>bol a du</i>
/ \	/ \
<i>aspect voice mood</i>	<i>person tense number</i>

All these verbal categories are semantic and functional categories, but not the categories of only grammar or lexis, such complicated categories are synthesized work of grammar, lexis and phonetics in syntax. In this continuum of expression, each category has its own element. For example, in the analytical predicate *okyp boladu*, the infix **-a-** is the indicator of Future Tense, the postfix **-du** is the indicator of person and number in auxiliary verb *bol*, but these elements are insufficient to express them separately.

The aspect, voice, and mood categories inherited from verb by gerund are in the root morpheme, but they are neutral in gerunds, when they function as the component of analytical predicates. Their properties are actualized and identified by auxiliary verbs, for example, the combination of *okyp boldu (had read)* is the continuum of expression of all verb category meanings, including the **result** aspect. The combination of *okyp zhatur (is reading)* expresses all the verb category meanings, including **the process** aspect.

Okyp boladu is the plane of expression of analytical predicate, the main verb *oky* transforms into

gerund **okyp**, the verb **bol** is grammaticalized as the auxiliary verb by neutralizing its lexical meaning, and its affixes **-a**, **-du** (**boladu**) express person, tense and number categories. The combination namely of this auxiliary verb with gerund expresses all verb category meanings, including the *result* action. Let us analyze the categorical verb meanings in this analytical continuums of expression *okyp boladu*:

- aspect: result;
- tense: future;
- voice: active;
- mood: indicative;
- number : single;
- person: III .

The Kazakh language has the plan of expression of these verbal categorical meanings synthetically in spite of that the synthetical plane of predicate expression is considered to be expressed by non-finites (gerund, participle), but not by finite verbs, only certain verbs (*tur, otur, zhur* and others and they are not so many) are considered to express verbal predicates synthetically.

The distribution of verb category meanings in the synthetical verb predicate *okidu*, which consists of the root morpheme *ok* plus affixation (*-i,-du*), is:

ok	i	du
/ \		/ \
aspect	voice	mood
	tense	number
		person

This verb predicate expresses all verb category meanings itself. In synthetical predicate *ok-i-du*, **-i** is an indicator of Future Tense, **-du** is an indicator of person and number, but these elements are not sufficient to express them separately without the verb *okidu*. All verb categorical meanings in a single continuum of expression *okidu* manifest themselves simultaneously :

- aspect: fact;
- tense: future, present (one form in **fact aspect** expresses two tenses: future and pseudo present);
- voice: active;
- mood: indicative;
- number : single;
- person: III .

Six different pieces of information are provided in one synthetical verb predicate **okidu** and analytical predicate **okyp boldu**. We can only be startled by the capacity of the human brain, which can decode these six integrated bits of information in a second. The fact that the *discrete meanings* (DM) of the six categories contained in these expressions are expressed simultaneously makes it difficult to tease apart the meaning of each category separately.


Sometimes this can lead to confusion; this is the reason for the existence of the traditional definition of aspect, which states that aspect is ‘*action occurs at the time or the distribution of the action in time.*’ Such temporal interpretation of aspect was mentioned by E. Koshmider, that aspect is ‘*time correlation from past to future*’ [3], on B. Comrie’s opinion, aspects are different ways of ‘*the internal temporal constituency of a situation*’ [4]; according to G. Guillante, it is the ‘*internal time of an action*’ [5]. E. Referovskaya offered that aspect is ‘*any action that is long or short contains in itself some quantity of ‘operative time’ of action*’ [6], ‘*aspect comes in a wider functional areas related to the ideas of time*’ [7].

According to this traditional definition of the aspect category, it has no independent meaning, because it

is created by the Tense category. Such a conflation of tense and aspect fails to account for the independent DM of each of these verb categories in general and aspect meaning particularly. This definition directly or indirectly influenced the opinion that some languages are temporal and some languages are aspectual. This definition led to conceptual and terminological confusion; this is the reason for the existence of the ‘joint’ tense-aspect category (TAC). If we accept this definition of the aspect category and joined nature of TAC, we open the door to denying the existence of DMs for all verb categories’ [2].

Time doesn’t generate any energy which is necessary for inner development of actions that is why Tense does not create actions. Where is no energy; there is no action.

Kazakh verb *seme* can generate these six verb category meanings. It is the first requirement of having of the aspect category. The second requirement of having of the aspect category is to have the planes of aspect expression. It has synthetical and analytical continuum of expression of all verb category meanings, including the aspect category. The Kazakh language has both requirements. If Kazakh verb generates six discrete meanings and has these meanings in its matrix and it manifests their meanings simultaneously by mentioned continuum of expression, can we say that some languages are aspectual and some languages are temporal?

The third objective reason for having of the verb aspect category is to have the plane of contents of the aspect category. Each notion in order to be a notion should have a plane of content and a plane of expression. The plane of content consists of concepts. The plane of content is primary, because it is a unit of thinking process. Conceptual identical feature characteristics of concepts are constant, unchangeable [7]. Concept ‘a tree’ is a tree everywhere. Probably there are some differences in configuration of leaves, colour, but everywhere it consists of leaves, a root, a trunk, and a crown. In spite of the concept  is pronounced differently in different languages, for example, *agash* in the Kazakh language, *derevo* in the Russian language, and *tree* in the English language, thanks to a unified plan of concept content, people from different nationalities understand each other.

The plane of the content of the aspect category is universal for all speakers, including Kazakh speaking people, it means that they have the plane of aspect content in their World View. What is the plane of content of the aspect category? It is inner stages of development of action as previously mentioned. Let us analyze the action: **to build** (a house). This action needs not only an intention of a person to build something, it also needs materials, human resources and finance for paying and buying. If a person intends to build something and all these enumerated items, which are necessary for performing of this action are provided, in that case this action has three inner stages of development. They are: **process action** (in order to complete an action, the process action should begin: the house is *being built*), then the **result action** where the action has been completed (the house *has been built*), after that it would be the **fact action**, *which states action* (the house *was built*).

If necessary materials are insufficient for completion of this action and they are not provided in time, in that case **protracted** action will take place. In that case this action ‘to build’ has four stages of development: **process**, **protracted**, **result**, and **fact**. Sometimes the necessary conditions would not be ensured for a long time or at all, in that case an action has only two stages: **process** and **protracted** actions.

And these four inner stages of development of ‘action’ have four planes of expressions as called the aspect. The plane of expression is secondary; it is an interpretive and decoding part of concepts, and it is changeable, how many languages so many pronunciation of this concept. We have never created concepts on the basis of a changeable plane of expression.

The Kazakh language has these four planes of content and four planes of expression of all verb category meanings, including the aspect category.

Types of Internal Development of Action (PC) and their Plane of Expressions	
PERFECT	
The plane of content	expresses a complete action. Action in its development comes to logical end of development and reaches a completely new stage.
The plane of expression	- gerunds (-п,-ып, -іп, -е,-а, -й) - the auxiliary verbs: біт, бол, кет, қал, ал, қой, сал, таста, шық, өт, жібер (bol, bit, ket, kal, al, koi, sal, tasta, shyk, ot, zhiber) + affixes of person, number, and tense categories.
PROCESS	
The plane of content	expresses a process action. Process is a main stage of actions, which is necessary for internal work to make prerequisites for completion and transit to a new stage.
The plane of expression	- gerunds -п,(-р), -а, е,-й. Айтү (verb)–aitup (gerund) - auxiliary verbs: тұр, отыр, жатыр, жүр + affixes of person, number, and tense
PROTRACTED ACTION AND ASPECT	
The plane of content	expresses a perdurative action which starts, but not completed for some reasons (insufficient conditions)and this protracted action is sometimes interrupted by other actions and it has some results before interrupting and it will be continued.
The plane of expression	-gerund with the indicators of gerund: -(п,-ып, -іп, -е,-а, -й) -the auxiliary verbs: тұр, отыр, жатыр, жүр, кел + affixes of person, number, and tense . -special adverbs (markers): әлі, содан бері, күні бойы, біраз, N сағат (бір сағат, 10 сағат; бір күн, он күн, 20-30 күн; үш жыл и др). Мен 3 сағат бойы тосып отырмын – I have been waiting for 3 hours.
FAST ACTION AND ASPECT	
The plane of content	expresses habitual, holistic, and automated actions. It includes systematic, successive repetition, axiomatic, momentary and other types of actions.
The plane of expression	Verb + affixes of the Present Tense:-й, -а, -е; -dy, -di, -ty, -ti. Verb + affixes -ды (du), -ді (di), -ты (ty), -ті (ti), which express general Fact actions in the Past. For example: bardu. Verb + affixes + -у(u), -тын (tyn),- тін (tin) + marker- еді, which express Iterated Past Fact Aspect. For example: baratyn Verb+ affixes –ған (gan), -ген (gen), -қан (kan), -кен (ken) + еді, which express the Past Fact actions happened for a long time ago. For example: bargan Verb + affixes + -а,-е,-й-айын (ain), - ейін (ein), - ын (yn), -ин express resolute future actions. For example: For example: barain. Verb + affixes +-мақ (maq), -мек (mek), -пақ (pak), -пек (pek), -бақ (bak), -бек(bek) express Intentional Future Fact actions, that an agent intends to do something in future. For example: barmakshu. Verb + affixes –ар (ar), -ер (er), -с(s) express possible future actions, the result of such actions are uncertain. For example: barar (maybe yes, maybe no)

Table 1. Types of Internal Development of Action (PC) and their Plane of Expressions

2.1 Subaspects

Subaspects is a further modification of the main aspects. They express themselves by the plane of the aspect expression, which they belong. Each main aspect has several sub-aspects. There are given only sub-aspects of perfect aspect.

2.2 Perfect Sub-Aspects

1. Iterative Result Sub-aspect expresses repeated result actions.

Ол Англияға екі рет барып келді. *He has been twice in England.*

2. Multistage Result or Simulfactive Sub-aspects expresses result actions on the basis of experiences.

Көпшіліктің ойы бойынша ұшақ жуық арада табылған жаңалық. Дегенмен ұшақтың осындай дамуы осы елу жыл ішінде жетіліп шықты.

Many people think that the rocket is a recent invention. Although the rocket has received its greatest development within the last fifty years.

3. Taxis Result Sub-aspects expresses several actions where one of these actions has finished before other .

Ол жұмысын ол келгенге дейін істеп бітірді. *He had finished his work before she returned.*

4. Instantaneous Result Sub-aspect expresses result actions occur fast.

Кенет жаңбыр жауып кетті. *It has rained suddenly.*

5 Suppositional Result Sub-aspects expresses actions which are supposed to have results.

Сіз бұл туралы оқып шығарсыз. *You will have read about it.*

7. Modal Result Sub-aspects expresses intensions

Оған мына жұмысты істеп кету керек еді. *He should have done this work.*

Process and Fact aspects have 8 sub-aspects and Protracted aspect has four sub-aspects, they have more that they have here. They are open for investigation.

3. Three Subjective Reasons Contributing to Deny the Existence of the Aspect Category in the Kazakh Language

The predicates of the Kazakh language are considered not to be expressed by the verbs synthetically, that they are expressed by non-finites (gerund and participle), it was previously mentioned. 'The main task of verbs is to express predicates. Verbs express predicates, because they can generate six verb categories, disclosing all action properties. Non - finites can not express predicates, because they don't inherit all verb categories, which are necessary for being a predicate, they have only some of them: aspect, voice and mood, which are insufficient to express predicates independently. Besides it, non- finites consist of verbs and other non-procession parts of speech as noun, adjective, adverb, which prevent to create verb action predicates. These two reasons do not allow non-finites to express predicates synthetically. However, they can participate in analytical predicates as one of the components of them together with auxiliary verbs. As it was previously mentioned, the auxiliary verbs actualize and identify the neutral verb categories of the gerund inherited from verb as aspect, voice, and mood and express the meanings of missed verb categories of the verb as tense, person, and number. The gerund and auxiliary verbs supplementing each other from the point of verb category meanings, can express verbal predicate.'

Why are verb predicates considered to be expressed by non-finites synthetically in Kazakh language? ‘Such infixes as *-a-*, *-i-*, *-e-* are considered to be the indicators of *gerund* (*aita- oilai, kore*). But they are the indicators of Future Tense in the plane of expression of verb synthetical predicates, for instance, *aitadu, oilaidu, koredi*. If non-finites are derived from verbs, why are these infixes considered to be only the indicators of *gerund*? According to the law of synthetic languages, each affix should have only one function. Having the polysemantic feature and the principle of language saving, it is strange to have such a law in language’ [2]. Any Kazakh verb expresses a verb predicate. It is the main task of verbs of any language.

The second reason of denying of the aspect category in the Kazakh language connects with Aktionsart. Being synthetic languages, the Russian language has Aktionsart, the Kazakh language has no it. As synthetic languages both of them have such notions as prefix, infix, postfix, and inflexion. But their function is different in these two languages, because the Kazakh language is an agglutinative part of synthetic languages, the Russian language is flexional. There is no pure synthetic, analytical, or agglutinative languages, all these structures are mixed in languages, the difference is that some languages are more or less synthetic or analytical than others. The Russian language is more synthetic than the Kazakh language.

The *infixes* of the Russian language in comparison with the Kazakh language are productive especially in organization of perfect (*reshIt*) and non-perfect (*reshAt*) aspects, on the basis of such infixes, the Slavic aspect theory has such notions as perfect and non-perfect aspects, though the number of aspects are more than two. Kazakh infixes don’t organize such aspect forms, because they have no such quality to express perfect and non-perfect aspects, they are the indicators of the tense category. It is also one of the reasons of denying of the aspect category in the Kazakh language.

The prefixes of the Russian language take part in organization of word formation, especially new verbs from verbs. For example: *отшуметь-otshumet*, *пошуметь-poshumet*, *зашуметь-zashumet*, *нашуметь-nashumet*. These verbs derived from the verb *шуметь-shumet* – ‘to noise’ with the help of different prefixes and denote different semantic variation of the verb *shumet*: *отшуметь* has the meaning ‘**to stop to make noise**’, *пошуметь* – ‘**to make noise for a while**’, *зашуметь* – ‘**to begin to make noise**’, ‘**нашуметь**’ - ‘**to noise enough**’ and they additionally express the Perfect aspect. The prefixes : *за-*, *по-*, *от-*, *на-* organize new verbs from initial verb root morpheme and denote the Perfect aspect.

These prefixes express indirectly the perfect aspect which are called as Aktionsart: *shumely* (non-perfect), *otshumely* (perfect), but they express different meanings, *shumely*- noised, *otshumely* – had stoped to make noise. Aktionsart is aspect within aspect. Modification of one verb is called differently, for instance the verb ‘*write*,’ if it expresses a process action (*is writing*), it is an aspect, if it expresses a perfect action (*have written*) it is an Aktionsart. On the basis of the plane of expression, the Russian language has ‘two’ types of aspects: Aspect and Aktionsart. Though new concepts are not created on the basis of the plane of language expression. If the plane of the aspect category is formed with the help of infixes (*reshIt-reshAt*) and suppletion (*idti-shol –go –went*), they belong to aspect. If the planes of aspect category are formed with prefixes, they are called Aktionsart.

Initially Aktionsart was considered to express the Perfect aspect, now Aktionsart ‘expresses’ non-Perfect aspect also, because verbs with prefixes can function as independent verbs, which can express all verb category meanings including perfect and non-perfect aspects, as independent verbs, the derived verbs from existing verbs can combine with infixes and can organize perfect and non-perfect aspects. For example, *pereraspredelil*- have redistributed (perfect) and *pereraspredelyal* - was redistributing/redistributed (non-perfect). As independent verbs they express all six verb category meanings, for example: *pereraspredelil*.

- aspect: **result**;
- tense: past;
- voice: active;
- mood: indicative;

– number : single;

– person: III

The aspect category has perfect and non-perfect aspects, so has Aktionsart. If the aspect expresses all development of inner stages of an action (*aity-aityp zhatur, aityp boldu, aitadu; to write - is writing, has written, has been writing*), so does Aktionsart (*pereraspredelil –pereraspredelal*).

Firstly, Russian prefixes are means of word-formation, in many cases they are verb-formation prefixes and newly organized verbs with prefixes are independent verbs, because they have the plane of content and the plane of expression. Verbs with prefixes or Aktionsart are not the question of the aspect category, they are the subjects of word derivation and derivation semasiology. In aspect theory, Aktionsart performs an uncharacteristic function for the aspect as explanation of meanings of verbs with prefixes. Aktionsart is not a language notion, it belongs to the concept ‘action’, it modifies inner development of an action, on the basis of person’s intention and provided conditions for performing of actions, as previously mentioned.

‘The phenomenon of Aktionsart has not been thoroughly studied in linguistics, and it still has no its clear status in the Aspect theory, either in and of itself or in term of its relation to the category of aspect’ [2]. However, this phenomenon is still used in aspectology. In A.V. Bondarko’s opinion, ‘notwithstanding the elaborate history of the study of ‘Aktionsart’, it should be noted that, in spite of the long and frequent usage of this terminology in aspectology, it is difficult to reach an understanding of it. ‘Aktionsart’ connects with phenomena concerning **how an action takes place and how that action is realized in time**’ [8]. In a wide sense, this idea of Aktionsart pertains to the definition of *aspect*, which states also that aspect is ‘*action occurs at the time or the distribution of the action in time*’. It is impossible when two ‘concepts’ (if the Aktionsart is the language concept) have one plane of content, one of them is created artificially or does not exist .

The Russian ‘Aktionsart’ also was the reason of denying of the existence of Kazakh aspect category, because it has no such aspect/verb formation prefixes as in Russian, in spite of the Kazakh language also belongs to synthetic language.

The third reason is the wrong definition of the aspect category, that was previously mentioned.

The absence of aspects were filled by tense category. There are about 30 tenses instead of three. They have so eclectic plan of contents, in spite of they are called as ‘tenses’, they express aspect, tense, mood, and modal meanings.

Five from six Kazakh Future tenses and three of Past Tenses express modality. It is seen from their names:

- *Сенімді келер шақ* (Confident (or: Assured) Future Tense)
Kazakh Grammar/Kazakh Academy of the SSA [1967:200]
- *Болжалды келер шақ* (Suppositional Future Tense)
Isaev, С. [2007:88][9], Ibragimov, К.[2011:59][10], Salkynbai, А. [2008:251] [11]
- *Мақсатты келер шақ* (Intended Future Tense)
Akhanova, D., Aldasheva, A., Akhmetzhanova, Z., Kadasheva, K., Suleimenova, E. [2002:111] [12]
- *Кесімді келер шақ* (Compulsary Future Tense) Zhubanov, К. [1999:387] [13]
- *Тілекті келер шақ* (Desired Future Tense) Zhubanov, К. [1999:388]
- *Күмән өнді өткен шақ* (Doubted Past Tense) Zhubanov, К. [1999:387]
- *Қатысты өткен шақ* (Participation Past Tense) Mamanov, И.Е. [1966][14]
- *Опық баяғылық өткен шақ* (Disappointed Past Tense) Zhubanov, К.[1999:387]

Some ‘tenses’ with one plan of expression represent two different plans of contents:

- Really Present Tense expresses (1) common and (2) long-term actions that can be observed at a given moment and during long time.

- General Present expresses the action which is happening right now, or probably, it is a protracted action.

Some planes of contents of these ‘Tenses’ are difficult to identify:

- Distant Past Tense expresses a distant past action and a speaker was not involved and unaware of this action.

- Non-Participation Distant Past Tense expresses that a speaker claimed that he did not participate in perfect and imperfect actions in the distant past.

Language is an objective phenomenon, it doesn't suffer from human beings' wrong judgments, learners suffer from it. Aspect is one of the main and dominant categories of a verb, defining propositions of sentences. Consequences of absence of the category of aspect in the Kazakh language we see from this sentence.

Қазақ хандығының төңірегіндегі мемлекеттер бұл уақытта күшейіп алған. Оларға түбі күшінін келмейтініне Әбілқайырдың көзі жете бастады (І. Есенберлин).

Countries around Kazakh Khanate **were gaining** strength and he realized that it would be beyond his power to control them.

The predicate of the original has the formula of the result aspect: **-ip + al -күшейіп алған** (*kusheip algan- had gained* strength). But the translation of the predicate expresses the process aspect: **were gaining**.

The original and translated version have different semantics on the basis of different predicates expressed by different aspects. According to the original version, khan knew that these countries could attack at any time, because they were the developed countries and the khan would do some actions to prevent it. On the basis of the translated version the khan had time to do something against these developing countries, he thought, that he had time to do something and he could be taken unawares by such strategy.

The idea of the author was broken from the point of logic proposition. The name of the khan is not translated, it was substituted by the pronoun ‘he’, but it does not make such semantic difference as aspect. The aspect category is needed for expressing the important nuances of actions. If concept is the essence of logical judgments, aspect is the essence of sentence proposition.

If translator does not catch thoroughly the proposition of a sentence, he can sometimes guesses or be guided by aspect formulas. If the Kazakh predicates have **-p-** plus one of these **auxiliary verbs** as **al, bol, bit** and others, it coincides to English formula: have + Participle II. It makes more easier our translation, communication processes.

4. Conclusion

1. The Kazakh language has the aspect category, because it has the plane of content, the plane of expression and Kazakh verb seme can generate all verb category meanings, including the aspect category.

Kazakh language has four main aspects: process, result, protracted, and fact.

These main aspects have several sub-aspects.

Kazakh aspects are expressed synthetically and analytically with the elements of agglutination.

Kazakh verbs have infixes –a,-e,-i in fact aspect, which express the Future Tense.

2. Traditional aspect definition does not disclose the notion ‘aspect’.

3. Aktionsart is an artificial invented notion on the basis of verb formation prefixes and it should belong

to the extra linguistic concept ‘action’, but not to language aspect. It is a subject of derivation or derivational semasiology.

Perspective: All Kazakh aspects and sub-aspects are open for further investigation from the point of their plan of content and the plane of expression.

References

1. Kakzhanova, F.A. (2015). *Категория аспекта в казахском языке и её подаспекты* [*The aspect category in Kazakh language and its sub-aspects*]. Karaganda. Monograph.
2. Kakzhanova, F.A. (2010). *Проблемы категории аспекта и её типология в английском языке*. [*The aspect category and its typology in the English Language*]. Karaganda. Monograph.
3. Koshmider, E. (1962). *Очерк науки о видах польского глагола. Опыт синтеза // Вопросы глагольного вида*. [*Sketch on Polish verb aspect. Experience of Synthesis // Questions about verb aspect*] М., 1962. С.105-167.
4. Comrie, B. (1976). *Aspect*. Cambridge Press.
5. Guillaume, G. (1973). *Immenceet transendance dans la categorie du verbe// Guillaume G. Langage et science duan langage*. 3rd ed. Paris: Niret. P. 47-48.
6. Maslov, Yu.S. (1994). *Очерки по аспектологии*. [*Sketches about aspectology*] Л. С. 1.
7. 20. Kakzhanova, F.A. (2013). *Views On Grammar*. California. FastPencil.
8. Bondarko, A.V. (2001). *Теория функциональной грамматики* [*Functional grammar theory*]. Moscow: UPPS. p63.
9. Isaevm, S.M. (2007). *Қазақ тілі: Оқу құралы* [*Kazakh language: Textbook*]. Алматы: Өнер, б.88.
10. Ibragimov, K. (2011). *Қазақ тілі: ҰБТ-ға дайындық оқулық тест* [*Kazakh Language: Preparation Tests for National Test Examination*]. Алматы. б.59.
11. Salkynbai, A. (2008). *Қазіргі қазақ тілі: Оқу құралы баспасы/* [*Kazakh Language: Textbook*], Алматы: Қазақ ун-ті. б.250.
12. Akhanova, D.H., Aldasheva, A.M., Akhmetzhanova, Z.K., Kadasheva, K.K., Suleimenova E.D. (2002). *Ресми-іскери қазақ тілі*. [*Official business Kazakh language*]. Almaty.
13. Zhubanov, K. (1999). *Қазақ тілі жөніндегі зерттеулер* [*Kazakh Language studies*]. Алматы: ҒЫЛЫМ, 1999. б.388.
14. Маманов, І.Е. (1966). *Қазіргі қазақ тілі* [*New Kazakh language*]. Алматы: Мектеп.
15. Kakzhanova, F.A. (2013). *Ағылшын тілінің функционалды грамматикасы* [*Functional English grammar*] for Kazakh speaking group. Almaty: Bastau. 307 б.
16. Kakzhanova, F.A. (1997). *Қазақ –ағылшын тілдерінің салыстырмалы грамматикасы* [*Kazakh English Comparative Grammar*]. Karaganda. 976.

The (non)realization of Armenian word-final palatal glides

Martin Macak
University of Georgia

Abstract

The (non)realization of Armenian word-final palatal glides is traditionally analyzed with reference to syllable-count and syntactic categorization: it is assumed that the glide is realized in monosyllabic nominals (e.g. *hay* [hɑj] ‘Armenian’; *goy* [gɔj] ‘existence’) but deletes elsewhere when in word-final position, i.e. in all polysyllables and verbs (e.g. *goy* [gɔ] ‘exists’; *ark^cay* [ɑr.k^hɑ] ‘king’). This analysis identifies the final nonrealized glide with specific morphological suffixes (*go-y* ‘exists’, i.e. ROOT-PRESENT.3SG.; *ark^ca-y*, i.e. ROOT-NOMINAL.ADJ, lit. “royal [one]”), whereas the final glide which is realized is either part of the root morpheme or the abstract noun suffix. A constraint-based optimality-theoretic analysis is proposed in which morphophonological constraints that enforce overt realization of specific morphemes (REALIZE MORPHEME) interact with constraints enforcing prosodic wellformedness (ONSET, NOCODA).

Keywords: Classical Armenian, Armenian orthography, phonology-morphology interface, optimality theory, prosodic wellformedness, syllable structure

1. Introduction

The realization of word-final -y [-j] is one of the few minor ways in which the traditional pronunciation of Classical Armenian (CA) departs from the original orthography designed to record the earliest attested Old Armenian (OA) language. The didactic rules of the classical orthography dictate that word-final -y [-j] be left unpronounced, see Table 1—with the exception of monosyllabic nouns, see Table 2.

TRADITIONAL ORTHOGRAPHY (CA. 400 AD)	TRADITIONAL PRONUNCIATION (CA. 1100+ AD)	GLOSS
<i>ark^cay</i>	[ɑr.k ^h ɑ]	‘king’
<i>k^cahanay</i>	[k ^h ɑ.hɑ.nɑ]	‘priest’
<i>k^cristoneay</i>	[k ^h ə.ris.tɔ.njɑ]	‘Christian’
<i>siroy</i>	[si.rɔ]	‘of/to/from love’
<i>beranoy</i>	[bɛ.rɑ.nɔ]	‘of/to/from mouth’
<i>Tigranay</i>	[tig.rɑ.nɑ]	‘of/to/from Tigran’
<i>stac^cay</i>	[əs.tɑ.ts ^h ɑ]	‘I obtained’
<i>goy</i>	[gɔ]	‘exists’
<i>kay</i>	[kɑ]	‘remains’

Table 1. Polysyllables and (monosyllabic) verbs with OA -y

<i>Hay</i>	[hɑj]	‘Armenian’
<i>xoy</i>	[χɔj]	‘ram’
<i>kay</i>	[kɑj]	‘station’
<i>goy</i>	[gɔj]	‘existence’
<i>bay</i>	[bɑj]	‘word, utterance’
<i>nay</i>	[nɑj]	‘humid, wet’
<i>vay</i>	[vɑj]	‘woe’
<i>ay</i>	[ɑj]	‘oh!, ho!’

Table 2: Monosyllabic nominals with OA -y

Crucially, the final palatal glide is traditionally pronounced in all polysyllables when covered by further morphology, i.e. before case endings and in composition, see Table 3. It is currently assumed that the word-final palatal glide deletes as a phonological segment in all polysyllables and monosyllabic verbs when in word-final position, cf., Godel (1975: 24); Schmitt (1981: 32); Vaux (1998: 20), etc.

<i>ark^cay-k^c</i>	[ɑr.k ^h ɑ(.)j(ə)k ^h]	‘kings’ (NOM.PL)
<i>ark^cay-s</i>	[ɑr.k ^h ɑ.jəs]	‘this king’ or ‘kings (ACC.PL)’
<i>ark^cay-ic^c</i>	[ɑr.k ^h ɑ.jits ^h]	‘of/to/from kings’
<i>ark^cay-owt^ciwn</i>	[ɑr.k ^h ɑ.ju.t ^h jun]	‘kingdom; reign’
<i>ark^cay-azn</i>	[ɑr.k ^h ɑ.jɑ.zən]	‘prince’ (‘royal-offspring’)
<i>ark^cay-a-kerp</i>	[ɑr.k ^h ɑ.jɑ.kɛrp]	‘royal, kingly, kinglike’

Table 3: Polysyllabic nominals with OA -y word-medially

2. Phonologically conditioned zero allomorphy

However, there is at least one monosyllabic nominal that falsifies the conventional hypothesis: *k^co-y* ‘of thy, yours (GEN.SG)’, which is traditionally pronounced [khɔ], not the expected *[k^hɔj]. In fact, this form has a later, hypercharacterized variant *k^co-y-oy* [k^hɔ.jɔ], i.e. /POSS.2SG-GEN.SG-GEN.SG/, which was arguably created to avoid ambiguity with two other homophonous forms: the possessive pronoun *k^co* [k^hɔ] ‘thy, your.NOM.ACC.SG’ and the GEN.SG form of the 2SG personal pronoun *k^co* [k^hɔ] ‘of thee’, which could be functionally differentiated from each other syntactically.

I argue that the pattern of (non)realization of the word-final palatal glide in the traditional pronunciation of CA may be explained by morphophonological principles that can refer to specific morphological units, rather than to syllable count and/or syntactic categories. The final palatal glide is always pronounced in lexical ROOTS (e.g. *xoy* ‘ram’ [χɔj] ← /xoi-/ ROOT-) or when it signals the ABSTRACT.NOUN SUFFIX /-i/ (e.g. *goy* ‘existence’ [gɔj] ← /gɔ-i-/). In contrast, when the final palatal glide represents other morphemes (see Table 4 below), it is only realized when parsed into syllable onset positions; cf. Table 3 above.

Glides affected by this pattern are always word-final, since glides are always realized when parsed into onsets, which is invariably the case before inflectional morphology. Since all morphology happens to start with a vowel, the affected suffixes are never parsed into codas in word-medial position, only word-finally; thus, *ark^ca y* ‘king’ OA *[ɑr.k^hɑj.] > CA [ɑr.k^hɑ] vs. *ark^ca-y-owt^ciwn* [ɑr.k^hɑ.ju.t^hjun] ‘kingdom’; *ark^ca-y-s* [ɑr.k^hɑ.jəs] ‘kings (ACC.PL)’ or ‘this king (NOM.SG)’, etc. The only exception is NOM.PL *ark^ca-y-k^c* ‘kings’, which is traditionally pronounced as [ɑr.k^hɑjk^h], i.e. with /-i/ NOMINAL.ADJ parsed into a word-final complex coda. This

realization, however, represents a later reduction of the conservative [ɑ.kʰɑ.jəkʰ] (← /ɑrkʰɑ-i-i-kʰ/ ROOT-ADJ-THEMEVOWEL-PL), which is preserved as such in the liturgical usage.

NON-REALIZED	E.G.,	MORPHOLOGICAL ANALYSIS	
GEN.DAT.ABL.SG	<i>beranoy</i>	/bɛran-ɔ-i/	‘of, to, from a/the mouth’
GEN.SG	<i>kcoy</i>	/kʰɔ-i/	‘of thy’ (ROOT-GEN.SG)
PRES.3SG	<i>goy</i>	/gɔ-i/	‘exists’
AOR.1SG	<i>keray</i>	/kɛr-ɑ-i/	‘I ate’ (ROOT-MED-AOR.1SG)
NOMINAL.ADJ	<i>arkcay</i>	/ɑrkʰɑ-i-/	‘king’ (lit. “royal [one]”)
REALIZED			
AOR.1SG	<i>Hay</i>	/hai-/	‘Armenian’
NOMINAL.ADJ	<i>bay</i>	/ba-i-/	‘utterance’
	<i>goy</i>	/gɔ-i-/	‘existence’
	<i>kay</i>	/ka-i-/	‘station; a standing’

Table 4: Nonrealized vs. realized OA -y

The assumption of a loss of specific suffixes but preservation of others partially explains the illusion of syntactic conditioning based on the split between verbs and nominals, cf. *go-y* ‘exist’-ABSTRACT.NOUN → OA/trad. [gɔj] ‘existence’ vs. *go-y* ‘exist’-PRES.IND.3SG — OA *[gɔj] ‘exists’ → trad. /gɔ-Ø/ [gɔ]. The other part of the explanation lies in the fact that the ABSTRACT.NOUN suffix */-i/, which regularly continues PIE */-ti/ ABSTRACT.NOUN, has been—by sheer historical coincidence—preserved only in synchronic monosyllables; cf., *bay* [bɑj] ‘utterance; speech; verb’ < PIE *bʰh₂-ti- (cf. Gk. φάσις ‘utterance’); *kay* [kɑj] ‘standing; station’ < *gʷh₂-ti- (perhaps cognate with Gk. βάσις ‘step; pedestal’); etc.

The diachronic explanation for such a pattern may be sought in the fact that the PIE abstract noun suffix */-ti-/ was a so-called primary suffix; i.e. */-ti-/ was selected by ROOTs (with restricted shapes), and this structural configuration was prototypically realized in two syllables. After the fixation of the pre-Armenian stress on the penultimate syllable and loss/reduction of the final syllable rimes/nuclei, the forms exhibiting this (synchronously) semi-productive suffix became invariably monosyllabic.¹

- **/CeH-ti-/ → PIE *bʰh₂-ti- (√bʰeh₂ ‘speak’) > *['ba.θi] > *['ba.ði] > *['ba.ji] > *bay* ['bɑj] ‘speech, discourse; word; verb’ (cf. Gk. φάσις ‘utterance; expression’)
- **/CeR-ti-/ → PIE *bʰr-ti- (√bʰer ‘carry’) > *['bar.ði] > *['bar.di] > *bard* ['bard] ‘pile; sheaf; (grammatical) compound’ (cf. Ved. *bhr-tī-ḥ* ‘support; bearing’)
- **/CeRC-ti-/ → PIE *wid-ti- (√weyd ‘find’) > *['γʷit.ti] > *['γiθ.ti] > *['giβ.ti] > *giwt* *['giwt] > traditionally pronounced ['gjut] ‘invention’ (cf. Skt. *á-vit-ti-ḥ* ‘not-finding’)
- **/CeC-ti-/ → *h₂d-ti- (√h₂ed ‘cut’) > ['hɑt.ti] > *['χɑθ.ti] > *['hɑβ.ti] > *hawt* *['hawt] > traditionally *hōt* [hɑt.] ‘division; flock’ (cf. Hitt. *ḫattāi-* ‘to cut’)
- **/CHeRC-ti-/ → *kh₂eyd-ti- (cf. Martirosyan 2010: 725) > *['khɑjd.ti] > *['kʰɑj.ti] (i.e. *COMPLEXCODA) > *['χɑj.tʰi] > *xaytʰ* ['χɑjʰtʰ] ‘sting; bite’ (cf. Lat. *caedo* ‘I hew, fell’)

¹ The PIE feminine ABSTRACT.NOUN suffix */-ti/ is in OA also reflected as part of complex suffixes, cf. */-ti-h₃on-/ > OA -*tʰiwn* (cf. Lat. -*tiōn-* ‘id.’) or the secondary */-ti/ abstract nouns formed on the bases of *s-stems, e.g. virtual */-ōs-ti-/ > OA -*owst* (e.g. *kor-owst* ‘loss; destruction’, cf. *kor-nč-i-m* ‘I am lost’). These suffixes became (highly) productive in OA, and so there is naturally no syllabic limit on their nominal formations.

The different treatment of specific */-i/ morphemes may be justified on functional grounds. The dropped OA inflectional suffixes (PRES.IND.3SG, AOR.IND.1SG and GEN.SG) either stood in opposition to the other endings within the relevant paradigms or were semantically expendable; cf. OA *k^cahana-y* *[k^ha.ha.naj] ‘priest-NOMINAL.ADJ’, i.e. ‘[a] priest ly [person]’ → CA [k^ha.ha.na-Ø] ‘id.’. The suffix that encoded nominal abstracts, however, did not stand in any paradigmatic or semantic opposition, and consequently its absence would have had serious semantic consequences; in other words, the very concept that the suffix indicates would have been irretrievably unexpressed. Similarly, if the final glide formed part of the root (cf. *hay* [haj] ← /hai-/), its surface realization dominated the ban on codas, presumably in order to functionally preserve lexical contrast.

In optimality-theoretic terms, a grammar that is sensitive to functional load and semantic recoverability of morphemes may be modelled by the promotion or demotion of constraints that enforce the realization of certain morphological elements (REALIZE MORPH) above or below constraints that enforce phonological markedness (e.g. ONSET and NOCODA).

When the morpheme consists of only one segment, the ranking PHONOLOGICAL MARKEDNESS » REALIZE MORPHEME virtually creates a specific type of allomorphy in which one of the allomorphs may be described as phonologically null. Thus, at some level of abstraction, we may morphologically describe the behavior of the suffixes that exhibit surface alternations as a phonologically conditioned zero allomorphy, i.e. */-i/ (≠ ABSTRACT.NOUN) → {-i} ~ {-Ø}.

The ABSTRACT.NOUN suffix was in the traditional pronunciation evidently not subject to deletion. I assume, therefore, that its realization dominated prosodic wellformedness principles (i.e., the constraint enforcing its realization was given priority over NOCODA, see section 4.5.). The illusion of the existence of a reference to monosyllabicity and nominal/verbal status may be therefore explained as a morphophonological process that may refer to or target specific morphological entities.

These assumptions may be tested by the traditional realization of *k^co-y* ‘of thy’, which is the GEN.SG form of the possessive pronoun *k^co* [k^hɔ] ‘thy; thine (NOM.ACC.SG)’. The traditional analysis predicts that as a monosyllabic nominal, *k^co-y* should be traditionally pronounced *[k^hɔj] as indicated by the traditional orthography;² however, *k^coy* is homophonous with *k^co* [k^hɔ] in the traditional pronunciation, and the two spellings are traditionally orthographically perfectly interchangeable; cf. *and k^co* ‘with thee’ (Bible+) ~ *and k^coy* ‘id.’ (e.g. Aragoncⁱ 1721).³

I assume that OA *k^co* diachronically reflects two inherited forms. As the suppletive GEN.SG of the personal pronoun *dow* ‘thou’ (i.e. in the meaning ‘of thee’), it continues an inherited **tweso* (cf. Hom. τέο, OCS česo ‘of what?’ < **k^weso*). As the NOM.ACC.SG form of the second person possessive pronoun, it may perhaps continue **two-s* (cf. Gk. σός, OCS *tvo-jb* ‘id.’). OA *k^coy*, the GEN.SG. of the latter *k^co* (< **twos*), continues an inner-Armenian **two-syo* and must have been synchronically segmentable into the lexical root morpheme *k^co-* ‘thy’ (< **two-*) characterized by the nominal GEN.SG case suffix *-y* (< **-syo*). The two forms, *k^co* ‘of thee; thy’ and *k^co-y* ‘of thy’, merged phonetically in the later language because the process in question specifically targeted the morphological ending, not monosyllabic nominals.

- *k^co* [k^hɔ] ‘thy, your (SG)’ NOM.SG.POSS.2SG < PIE **two-s* (cf., Gk. σός, OCS *tvo-jb* ‘id.’)
- *k^co* [k^hɔ] ‘of thee, of you (SG)’ GEN.SG.PRO.2SG < **tw-eso* (cf., Hom. τέο, OCS česo ‘of what?’ < PIE **k^w-eso*); cf. *dow* ‘thou’ NOM.SG.PRO.2SG < PIE **tu* (cf. E *thou* ‘id.’)
- *k^co-y* [k^hɔj] ‘of thy’ GEN.SG.POSS.2SG < OA *k^coy* *[k^hɔj] < QIE **two-syo* (cf. Skt. *tásya* ‘of that/him’ < PIE **to-syo*)

² Additionally, we may assume that the final palatal glide was originally pronounced based on the shape of the hypercharacterized variant *k^coy-oy* ‘id.’, which preserves the underlying palatal segment in word-medial position (i.e., *k^coy/k^ho-i/* ‘thy-GEN.SG’ → *k^co-y-oy* ‘id.-THEMEVOWEL-GEN.SG’).

³ The canonical distribution of these forms in OA is nicely illustrated in Judith 11:4; *Ənkál zbans alaxnoy k^coy* (GEN.SG of *k^co* ‘thy’), *ew xōsescⁱ alaxin k^co* (POSSESSIVE.2SG.NOM) *arají k^co* (GEN.SG of *dow* ‘thou’) “Receive the words of thy servant and suffer thine handmaid to speak in front of thee.”

Furthermore, this original morphophonological pattern is in the medieval period overapplied and leads to the spelling convention of ‘covering’ any word-final vowel with the silent grapheme for the palatal glide, regardless of syllable-count or syntactic category. Thus, written final sequences such as *°ay* may in the post CA text or even in the extant (redacted) copies of OA texts stand for an original OA *°a*; i.e., final orthographic *°a* and *°ay* are practically interchangeable—even in monosyllabic nominals! For instance, many versions of the anonymous Armenian translations of the *Art of Grammar* by Dionysios Thrax write the names of letters such as ԴD (= OA *da*) as *day* (pronounced [da]), ՋZ (= OA *za*) as *zay* [za], etc (cf. Adontz 1970: *passim*).

A grammar responsible for the morphophonological pattern just described will be formalized using Optimality Theory (Prince and Smolensky 2004) in (5.5). In it, I assume that morphophonological constraints that enforce realization of specific morphemes (REALIZE MORPH) interact (by ranking) with constraints enforcing prosodic wellformedness, in this case, at the level of syllabic structure (ONSET and NOCODA).

3. Functional dispensability and syllabic markedness

If we take the traditional analysis at its face value, it actually assumes that a specific phonetic segment is phonologically different when it builds word-forms functioning as one class of lexemes than when it is used to build word-forms functioning as another lexical category: the final glide is thus presumably preserved in *Hay* [hɑj] ‘Armenian’ because the lexeme is a noun but presumably deleted in *kay* *[kɑj] (> trad. [kɑ]) because the lexeme is a verb. As we have seen in the previous section, however, the reference to lexical categories is arguably epiphenomenal. The final segment that is either covert/suppressed or overt/expressed forms part of independent morphological suffixes (with specific semantic value). If we look at the value of the suffixes that are dropped from a functional angle, the suffixes appear to be either morphologically recoverable or semantically dispensable.

Let us look at the final *-y*’s in CA *verbs* regardless of their surface syllabic count. These glides are invariably suffixed inflectional morphemes, namely PRESENT.INDICATIVE.3SG and AORIST.INDIC.MEDPASS.1SG. The paradigms in (1) show a simple morphological analysis of the present indicative of the OA verb *la-m* ‘I weep’ and its continuation (with a modified syntactic function) into modern Armenian.

(1)	OA PRES.IND	EA FUT.SUBJ	WA PRES.SUBJ
√-1SG	<i>la-m</i>	<i>la-m</i>	<i>la-m</i>
√-2SG	<i>la-s</i>	<i>la-s</i>	<i>la-s</i>
√-3SG	<i>la-y</i> *[lɑj]	<i>la-Ø</i> [lɑ]	<i>lay</i> [lɑ] = /lɑ-Ø/
√-1PL	<i>la-mk^c</i>	<i>la-nk^c</i>	<i>la-nk^c</i>
√-2PL	<i>la-yk^c</i>	<i>la-k^c</i>	<i>la-k^c</i>
√-3PL	<i>la-n</i>	<i>la-n</i>	<i>la-n</i>

The final *-y* in WA *lay* is purely orthographic; the form is pronounced [lɑ], and the simplest morphological analysis arguably contradicts the traditional orthography. Observe that the loss of the 3SG *°-i/* suffix has no consequences for the functional contrast of word-forms within the paradigm. This is perhaps one of the reasons losses of inflectional suffixes are typologically quite common. I argue that the simplest analysis of the data in (1) is to assume that the 3SG ending *°-i/* (→ *°[-j]*) was lost (or replaced by *°-Ø/*) on its way to the modern language, just as 2PL *°-yk^h/* was obviously replaced by *°-k^h/*.⁴

The same analysis may be assumed in the case of the aorist first person singular ending */y/* *°[-j]* and the genitive (dative, ablative, and locative) singular ending *°-y/* *°[-j]* on polysyllabic nominals such as OA *ker-a-y*

⁴ The change of CA 1PL *°-m(-)k^h/* (= *°[-møk^h]* > *°[-mk^h]*) to modern Armenian *°-nk^h/* *[-ŋk^h]* is apparently regular.

*[kɛrɑj] ‘I ate’ or OA *Tigran-a-y* *[tigrɑnɑj] or OA *beran-o-y* *[bɛrɑnɔj] ‘mouth-THEM.V-GEN.SG.’, traditionally pronounced [kɛrɑ], [tigrɑnɑ] and [bɛrɑnɔ], respectively. These various */-i/ suffixes were on the way to the modern language subject to prosodic well-formedness when its surface reflex was to be parsed into a coda after a vowel. Instances of the GEN.SG */-i/ had been evidently subject to prosodic well-formedness already in pre-OA. In OA, the GEN.SG desinence of the *i*-stem inflection is -ḫ -*i*, not the expected **-i-y*, which synchronically represents */i-j(ɔ)/ (i.e. -THEM.VOWEL-GEN.SG). The palatal glide was lost in this desinence already in Pre-OA due to an unrelated phonological process (the OBLIGATORY CONTOUR PRINCIPLE; cf. also INST.SG. -m -ow [-u] < *[-u(w)] < PA */-u-βj/ < PIE *-b^hi) in which the surface realization of the morpheme merged with the realization of the homorganic thematic vowel into the attested OA *-i* [-i].

4. Polysyllabic nominals with the semantically expendable nominalizer */-i/

The argument that the final *-y* on polysyllabic nouns is to be interpreted as a synchronically recognizable suffix is provided by the phonological development of original final syllables in the relatively recent pre history of OA. The synchronic final stress is historically a result of the original penultimate stress prior to the loss of the original final rimes, e.g. *e-ber* [jɛ.ˈbɛr] ‘carried’ < *[ɛ.ˈbɛ.rɛ] < *é-b^her-e-t (Ved. *ábharat*, Gk. ἔφερε).⁵

Early loanwords from Syriac into pre-OA also exhibit loss of final syllables; e.g., Syriac *lēqā* ‘rudder’ ultimately becomes a monosyllabic OA *lek* ‘id.’, see (3). However, the original final Syriac syllable is preserved if a lexeme was extended with the adjectival suffix */-i-/ before the loss of final syllables; e.g., Syriac *kāhānā* ‘priest’ ultimately becomes OA *kāhana-y* ‘id.’ (not ***kāhan*) via */k^hahana+i i-/, see (3). The lists in (2) and (3) are taken from Biblical evidence provided by Olsen (1999: 931ff.).

(2) Syriac	OA	Gloss
<i>gezzəθā</i>	<i>gzat^c</i>	wool
<i>gubbā</i>	<i>gowb</i>	hole; cistern
<i>dārā</i>	<i>dar</i>	age; generation
<i>zauḡā</i> (from Gk. ζεύχος)	<i>zoyg-k^c</i>	pair; couple
<i>targmānā</i>	<i>t^cargman</i>	interpreter
<i>xargəlā</i>	<i>xaragowl</i>	type of locust
<i>ṣaumā</i>	<i>com</i>	fast(ing)
<i>ṣərārā</i>	<i>crar</i>	bundle
<i>qaqqəwā</i> (cf. Gk. κაკκαβη)	<i>kak^caw</i>	partridge
<i>qəlāfā</i>	<i>kelew</i>	cortex, skin
<i>xawlā</i>	<i>haḫb-k^c</i>	snare; net
<i>lēqā</i>	<i>lek</i>	oar; rudder
<i>məgalləθā</i>	<i>magalat^c</i>	parchment
<i>mezzē</i>	<i>maz</i>	hair
<i>maggəlā</i>	<i>mangaḫ</i>	sickle
<i>meškā</i>	<i>mašk</i>	skin
<i>maḫsā</i>	<i>mak^cs</i>	custom; tax

⁵ As far as the relative chronology of the apocope of final rhymes is concerned, there is enough evidence to assume that the inherited final syllables were still present in the language during the time of the earliest Iranian loanwords, as has been convincingly argued by Olsen (2005) and reflected in the earlier opinion of a number of scholars (e.g. Meillet 1911: 149, 1936: 23; Jensen 1959: 19; Olsen 1999: 859; Matzinger 2005: 27ff.).

<i>maṭrəqā</i>	<i>mtrak</i>	whip
<i>nəšifā</i>	<i>nših</i>	fine flour
<i>šabbəθā</i>	<i>šabat^c</i>	week; Sabbath
<i>šəwīlā</i>	<i>šawił</i>	path
<i>‘ərūwəθā</i>	<i>owrbat^c</i>	Friday
<i>şeşşā</i>	<i>c^cic^c</i>	peg, plug
<i>kakkərā</i>	<i>k^cank^car</i>	talent, mina
<i>kārōzā</i>	<i>k^caroz</i>	herald; preacher
<i>kərowā</i>	<i>k^cerob</i>	cherub
<i>kumrā</i>	<i>k^cowrm</i>	priest
<i>kewšā</i>	<i>k^cawš</i>	he-goat
(3) Syriac	OA	Gloss
<i>zōpā</i>	<i>zopa-y</i>	hyssop
<i>lūmā</i>	<i>lowma-y</i>	mite
<i>şişşəlā</i>	<i>cnčla-y</i>	cymbal
<i>kaskərā</i>	<i>kaskara-y-k^c</i>	grill (for cooking)
<i>mannā</i>	<i>manana-y</i>	manna
<i>sātānā</i>	<i>Satana-y</i>	Satan
<i>srīqā</i>	<i>srika-y</i>	murderer; assassin
<i>ṭalyā</i>	<i>ṭla-y</i>	child; boy
<i>filōsōfā</i> (< Gk. φιλόσοφος)	<i>p^cilisop^ca-y</i>	philosopher; scientist
<i>kāhānā</i>	<i>k^cahana-y</i>	priest
<i>quwyā</i> (< Gk. κυβεία)	<i>k^cowa-y</i>	cube; divination die

The appended nominalized adjective suffix *-y* apparently refers to either *the characteristic property of a person* (profession, ethnicity, social status) or *the inherent characteristics of things or materials*; cf. *srika-y* (< Syr. *srīqā* ‘murderer’ + */-i-/) which also means ‘ruffian, rascal, hooligan’ literally ‘[a] murderous, criminal [person]’; *ṭla-y* (< Syr. *ṭalyā* ‘boy’ + */-i-/) literally ‘[a] boyish, young [person]’ (cf. Lat. *adolēscēns* ‘(adj.) growing up, maturing; (subst.) a mature person’; *Satana-y* (< Syr. *sātānā* + */ i /) “[the/a] devilish, adversative [being]; Satan”.⁶

5. Surface alternations resulting from positional faithfulness

When the morpheme consists of only one segment, as is the case with various homophonous */-i/’s, its overt non-realization may be functionally reanalyzed as a zero morpheme /-Ø/; e.g. /ka-i/ ‘remains’ ROOT-PRES.IND.3SG → *[kaj] > [ka] → /ka-Ø/. If, however, a morpheme’s overt (non-)realization is a function of phonological context, cf. /ark^ha-i/ → (OA *[ar.k^haj] >) CA [ar.k^ha] ~ /ark^ha-i-e/ ROOT-NOMINAL.ADJ-ABL. SG → [ar.k^ha.je], the process may be conceptualized as a phonologically-driven allomorphy involving a

⁶ Typological parallels may be found in Slavic languages, cf. Czech *služebná*, which is formally an adjective, e.g. *služebná zbraň* ‘service gun’, but which is also lexically a nominalized adjective, i.e. ‘[a woman] serving; service maid’. Abundant examples may be found in Romance or Germanic languages; cf. NE *Jewish* (referring to a person) and OA *Hr(-)ea-y* ‘Jew’, from Syriac *’ihūdāyā* ‘Jew, Hebrew’ + */-i-i-/ NOMINAL.ADJ (later actually reanalyzed as /hur-eay/), i.e. properly “[a] Jewish [person]”. ‘φ’ NOMINAL.ADJ “Characteristic Property of φ”.

conditioned zero allomorph, i.e. NOMINAL.ADJ */-i/ → {-Ø} ~ {-j}. The zero allomorph is selected should the morpheme end up in a syllable coda; the ‘overt’ allomorph surfaces in onset positions; cf. CA [ɑr.k^hɑ-Ø] ‘king’ vs. CA [ɑr.k^hɑ-.j-e] ‘from [a/the] king’.⁷

6. OT analysis

In optimality-theoretic terms, a grammar which is sensitive to functional load and semantic recoverability of morphemes may be modelled by the promotion or demotion of constraints which enforce the realization of certain morphological elements (REALIZE MORPH) above or below constraints which enforce phonological markedness (e.g. ONSET and NOCODA). The realization of phonological material which is part of a lexical root or of the ABSTRACTNOUN */-i/ (in incidental monosyllables) is modeled in Tableau 1 and Tableau 2, respectively. The constraint MAX-ROOT defined in (4) below enforces realization of all material on lexical roots.

(4) MAX-ROOT

“Do not delete any root segment” (de Lacy 2002; Yu 2007: 79)

/hai-Ø _{NOM.ACC.SG} / ⁸	MAX-ROOT	ONSET & NOCODA
→ a. .haj.		*
b. .ha.	*!	

Tableau 1: Preservation of stem-final /i/ in *Hay* ‘Armenian’

/gɔ-i _{ABSTRACT} /	MAX-ROOT	REALIZE /-i/ _{ABSTRACT}	ONSET & NOCODA
→ a. .gɔ-j.			*
b. .gɔ-Ø _{ABSTRACT}		*!	
b. .g-i.	*!		

Tableau 2: Realization of ABSTRACTNOUN */-i/ in *go-y* ‘existence’

When the morpheme consists of only one segment, the ranking PHONOLOGICAL MARKEDNESS » REALIZE MORPHEME virtually creates a specific type of allomorphy in which one of the allomorphs may be described as phonologically null. Thus, at some level of abstraction, we may morphologically describe the behavior of the NOMINAL.ADJ suffix */-i/ as a phonologically conditioned zero allomorphy, i.e. NOMINAL.ADJ */-i/ → {-i} ~ {-Ø}. Since the suffix */-i/ is phonologically suppressed only when potentially parsed into the coda, the constraint which penalizes the realization of the suffix in this case may be identified as NOCODA; see Tableau 3.

(6) ONSET

“All syllables have to have onsets.”

NOCODA

“All syllables must not have a coda.”

REALIZE MORPHEME

“For every morpheme in the input, some phonological element should be present in the output.”

⁷ The original realizations of the traditional [ɑr.k^hɑjk^h.] and [ɑr.k^hɑjs], i.e. NOM.PL and ACC.PL, respectively, was [ɑr.k^hɑ.jək^h] and [ɑr.k^hɑ.jəs], as is still observed in the liturgical usage. The forms reflected earlier */ɑrk^hɑ-i i-k^h/ ROOT-ADJ-THV-NOM.PL with penultimate accent and reduction of the final syllable nucleus, i.e. *[ɑr(k^hɑ.jək^h)] (cf. *ayžm* *[ɑj.ʒəm] ‘now’ from *[(ɑjs.ʒəm)] (lit. ‘this hour’, cf. Sp. *ahora* ‘id.’).

⁸ The underlying forms of all nominal elements (roots and suffixes) also contain specific thematic vowels which are apparently not realized on the surface unless stressed; e.g. /hai-ɔ-Ø_{NOM.ACC.SG}/ → [ˈhaj] : /hai-ɔ-ts^h_{GEN.DAT.ABL.PL}/ → [ha.ˈjots^h]. The derivation of all the forms analyzed here is in this respect simplified, but this approximation does not affect the spirit of the proposed argument.

/ark ^h a-i-Ø _{NOM.ACC.SG} /	MAX-ROOT	ONSET	NOCODA	REALIZE /-i/ _{ADJ}
→ a. .ar.k ^h a.-Ø _{ADJ} -Ø _{NA.SG}				*
b. .ar.k ^h a.-j.-Ø _{ADJ} -Ø _{NA.SG}			*!	
b. .ar.k ^h a.-i.-Ø _{ADJ} -Ø _{NA.SG}		*!		

Tableau 3: Non-realization of NOMINAL.ADJ */-i/ in codas

(van Oostendorp 2005)

The semantically expendable suffix in /ark^ha-i-Ø_{NOM.ACC.SG}/ potentially parsed into a syllable coda—as arguably in OA *ark^ca-y* *[ar.k^haj]—is optimally nonrealized (or morphologically realized by its covert /-Ø/ allomorph), if its overt realization would render this (see candidate b) and any potential output (see candidate c) sub-optimal. The suffix is, however, still recognizably present underlyingly. Since there are no phonological principles which would suppress the realizations of onsets, the suffix surfaces overtly realized if optimally parsed into the onset position, see Tableau 3 below.

/ark ^h a-i-ut ^h iun/ “royalty”	MAX-ROOT	ONSET	NOCODA	REALIZE /-i/ _{ADJ}
→ a. .ar.k ^h a.-Ø-.u.t ^h jun.		*!		*
b. .ar.k ^h a.-Ø-.w.t ^h jun.				*!
b. .ar.k ^h a.-j-.u.t ^h jun.				

Tableau 4: Realization of NOMINAL.ADJ */-i/ in onsets (TETU)

Just as in onsets, the proposed grammar does not penalize the realization of the suffix in syllabic nuclei either; e.g. *caraw* ‘thirst; draught’ → *caraw-i* ‘thirsty; dry’; *ał* ‘salt’ → *ał-i* ‘salty’; *t^cšnam-i* ‘of evil mind; hostile’, hence ‘enemy’ etc., see Tableau 5.

/ał-i- Ø _{NOM.ACC.SG} /	MAX-ROOT	ONSET	NOCODA	REALIZE /-i/ _{ADJ}
→ a. .a.ł-i.				
b. .a.ł.-Ø _{ADJ}			*!	*

Tableau 5: Realization of NOMINAL.ADJ */-i/ in nuclei

Along the same lines, we may conceptualize the historical loss of certain suffixes (such as the inflectional suffixes mentioned earlier, namely PRES.IND.3SG, AOR.IND.MED-PASS.1SG or GEN.SG) as a complete demotion of the morphophonological constraints that enforce their realization. Since in such cases, all sub-optimal output is harmonically bound, i.e. the endings are not part of any synchronic alternations—cf. [ark^ha-Ø] ‘king’ ROOT.NOMINALADJ.NACC.SG ~ [ark^ha j-ut^hjun] ‘kingdom’ ROOT.NOMINALADJ-ABSTRACTNOUN.NOM.ACC.SG—we may consider these suffixes simply historically lost (or replaced by a synchronic invariant zero /-Ø/). The loss of overt phonological material appears to be a direct result of the functional recoverability of the morphemes represented by this material.

7. Palatal glide realized in complex codas

The realization of the NOMINAL.ADJ /-i/ in forms such as *ark^ca-y-k^c* [ar.k^hajk^h.] ‘kings (NOM.PL)’, with the palatal glide in a complex word-final coda, continues an earlier pattern with the canonical [ar.k^ha.jk^h] (← /ark^ha-i-i-k^h/ ROOT-ADJ.-THV-PL), with the glide in the onset, penultimate accent and reduced post-tonic thematic vowels (cf. *ay-n* ‘this [one] NOM.ACC.SG’ : *ay-no-r* ‘id. GEN.SG’, i.e. //ai-no-Ø// : //ai-no-rV// → /'aj.nə/ : /aj.'nə.rə/ → ['ajn] : [aj.'nəɾ]). The form [ar.k^hajk^h.], the yet more innovative realization of the suffix in the complex coda, may be understood as the output of the grammar of later medieval and/or modern Armenian without underlying thematic vowels. This provision easily explains both variants; cf. Tableau 6 below.

/ark ^h α-i-k ^h / “royal [ones]”	REALIZE /-kh/PL	NoCODA	REALIZE /-i/ADJ	*COMPLEXCODA
a. .ar.k ^h α-j.-Ø	*!	*		
b. .ar.k ^h α.-Ø-Ø	*!		*	
c. .ar.k ^h α-Ø-k ^h .		*	*!	
→ d. .ar.k ^h α-j-k ^h .		*		*

Tableau 6: NOMINAL.ADJ /-i/ surfaces as part of a complex word-final coda (cf. TETU).

8. Conclusion

I have argued that the observed phenomenon is a morphophonological process in which realization of specific morphological categories is conditioned by well-formedness of syllabic structure, namely the preference for open final syllables. When a morpheme that violated this preference was functionally expendable—i.e., it was recoverable based on paradigmatic opposition (cf. *kay* /ka-i/ = /ka-Ø/ ‘stands’) or it was semantically redundant (cf. *k^cahanay* /k^hahana-i/ ‘[a] cleric-*al* [person] ≈ [k^hahana-Ø] ‘cleric-Ø’)—the phonological material of the morphological element was not overtly realized.

When, however, lexical contrast or semantic recoverability were to be compromised (cf. *Hay* ‘Armenian’, i.e. ROOT ≠ *[ha]; ka-y ‘standing, station’, i.e. ROOT-ABSTRACT.NOUN ≠ *[ka], etc.), the overt realization of the morpheme dominated syllabic markedness. Incidentally, the suffixes that were recoverable did not appear on monosyllabic nouns—hence the illusion of a reference to lexical categories. The relevant nominal morphemes were either roots with the unproductive ABSTRACT.NOUN suffix inherited in monosyllabic forms or lexical roots inherited as monosyllables—hence the illusion of a reference to syllable count.

References

- Adontz, N. (1970). *Denys de Thrace et les commentateurs arméniens*. Louvain: Impr. Orientaliste.
- Aragonc'i, P. (1721). *Girk^c aṙak^cinow^ceanc^c (De virtutibus)*. Venetik: tparan Antoni Bortalii.
- Godel, R. (1975). *An introduction to the study of Classical Armenian*. Wiesbaden: Reichert.
- Künzle, B. (1984). *Das Altarmenische Evangelium = L'Évangile arménien ancien*. Bern: Lang.
- de Lacy, P. (2002). Maximal Words and the Maori Passive. In Norvin Richards. ed. *Proceedings of the Austronesian Formal Linguistics Association (AFLA) VIII*. Cambridge, MA: MIT Working Papers in Linguistics 44. 20–3.
- Olsen, B. A. (1999). *The Noun in Biblical Armenian: Origin and Word-formation. With Special Emphasis on the Indo European Heritage*. Berlin: Mouton De Gruyter.
- Olsen, B. A. (2005). On Iranian Dialectal Diversity in Armenian. In Olav Hackstein and Gerhard Meiser (eds.). *Sprachkontakt und Sprachwandel: Akten der XI. Fachtagung der Indogermanischen Gesellschaft, 17.-23. September 2000, Halle an der Saale*. Wiesbaden: Reichert. 473-481.
- Prince, A. & Smolensky, P. (2004). *Optimality theory: constraint interaction in generative grammar*. Malden, MA: Blackwell Pub.
- van Oostendorp, M. (2005). Expressing inflection tonally. *Catalan Journal of Linguistics* 4(1). 107–127.
- Schmitt, R. (1981). *Grammatik Des Klassisch-Armenischen mit Sprachvergleichenden Erläuterungen*. Innsbruck: Institut für Sprachwissenschaft der Universität.
- Vaux, B. (1998). *The phonology of Armenian*. Oxford: Clarendon Press.
- Yu, A. C. L. (2007). *A natural history of infixation*. Oxford: Oxford University Press.
- Zhang, J. (2001). *The Effects of Duration and Sonority on Contour Tone Distribution—Typological Survey and Formal Analysis*. University of California.

A Verb *aa* in Dagur

Yohei Yamada

Tokyo University of Foreign Studies
Research Fellowship for Young Scientists of JSPS

Abstract

A verb *aa* in Dagur is used in three ways. (a) as an existential verb, (b) as a copula verb, or (c) as an auxiliary verb, the same as *bai* in Mongolian. My paper shows two points of usage of this verb: *aa* does not have any negative forms in all the manners (a)~(c), and it does not appear as the “simple form” with non-past verb final form in (a) or (b), but not in (c). If we are to express the present tense un-marked existense meaning, an adjectival expression with *bei* is employed. Such this usage of existential verb is superseded by **bayi* in the Northern group of Mongolic. In the South group of Mongolic, almost all cognates with **a* are used but they are gramaticalized and hence lose their verbal property. Existential verb *aa* in Dagur is rather similar to *a* in Middle Mongolian but not to other synchronic Mongolic languages.

Keywords: Dagur, existential verb, copular verb, auxiliary verb

1. Introduction

The verb *aa* in Dagur is used in three ways: (a) as an existential verb, (b) as a copular verb, and (c) as an auxiliary verb, all of which are translated as “to be” in English.

- (1) *šii udiš orie haane aa-sen=šie?*
2SG yesterday night where AA-PERF=2SG.Q
Where were you last night? ((a) existential) Shiotani (1991: 76)
- (2) *tednii neu-ǰ ire-gu-d-ini bii učiiken aa-sen=bie.*
3PL.GA to.move-SIM to.come-VN-DAT-3SG 1SG little AA-PERF=1SG
When they moved here, I was little. ((b) copula) Shiotani (1991: 59)
- (3) *bii šamii eri-iči-gu-d-mini, yoo kii-ǰ+aa-sen=šie?*
1SG 2SG.ACC to.seach-to.go-VN-DAT-1SG what to.do-SIM+AA-PERF=2SG.Q
What were you doing, when I visited you? ((c) auxiliary) Shiotani (1991: 76)

This verb is a cognate with the verb *a* in Middle Mongolian and with some fossil words such as *aǰ axui* “economics” and *atlaa* “however” in Mongolian.

This paper shows that some usage of the verb *aa* is different from other verbs and the differences clarify that the auxiliary use of *aa* is rather similar to a suffix. In comparison to other Mongolic languages, the verb *aa* in Dagur is closer with the cognate *a* of Middle Mongolian.

The use of the verb *aa* in three ways has been replaced with **bayi* in Central Mongolics and their existential verb **bayi* differs from *aa* in Dagur in that they have the same negative form as other verbs and that

they are often used with nonpast verb final forms in copular use. The cognates with *aa* in Dagur are found in Monguor and perhaps in Bonan, the South group of Mongolic; however, they function more as a particle than as a verb.

The section below provides a profile of Dagur, followed by a demonstration of the usage of the verb *aa*. Section 4 explains the differences between existential and copular use of *aa* and auxiliary use; section 5 compares the usage of *aa* in other Mongolics; section 6 provides a conclusion.

2. Profile of Dagur

Dagur (Daur, Daor, or Daguur) is one of the Mongolic languages, spoken mainly in the Northeast China. There are four dialects in Dagur: Butha, Qiqihar, Hailar and Xinjiang. Butha is spoken by the largest population group, living in Morii Dawaa Banner in Hulunbuir, Inner Mongolia. Butha Dagur is considered Dagur's standard dialect. Note that from my investigation, the usage differences of the verb *aa* in any dialects are subtle, so we do not care about the differences between dialects.

3. The verb *aa* in Dagur

3.1 Existential *aa*

The existential *aa* employs the past form *aa-sen* (4) and the progressive form *aa-ǰ+aa-wei*¹ (5), but it does not have a simple form. If we are to express the unmarked existential meaning in the present tense, such as “there is ...,” an adjectival expression with *bei* is employed (6)². *bei* is not a verb and does not have any verbal form such as the past tense. In a negative sentence, *uwei* is employed (7). *uwei* is similar to *bei* in that it does not have any verbal forms.

- (4) *ter arben doloo bol-tl-oo hoten-d aa-sen.*
 that ten seven to.become-LIM-REF town-DAT AA-PERF

He was in a town until he was 17 years old.

Shiotani (1991: 59)

- (5) *nek buduun čoloo aa-ǰ+aa-wei.*
 one big stone AA-SIM+AA-NPST

There is a big stone.

- (6) *baawaa-šini haane bei.*
 father-2SG where exist

Where is your father?

Shiotani (1991: 53)

¹ In my investigation of Hailar Dagur, non-past is attested as *-wei* for both first person singular and third person (though Enghebatu et al. (1988) describe *-wei* as non-past first person singular). In other dialects, *-bei* is often seen.

² The existential verb, which I call here, may be terminologically confusing because it is used both for existential construction and for locative construction. Fundamentally, these expressions in Dagur may not be distinguished each other by word order; a locative argument put after subject in (4) is interpreted as in locative construction; on the other hand, one put before the subject or not represented in (5)-(7) is in existential construction.

Other than this, *-tii* (<-tAi), possessive property meaning or adjectival suffix is also used for existential meaning. (e.g. *ene wairelčuu gajir baraan čoloo-tii* {this near place many stone-PROP} ‘There are many stones in this place’ (Shiotani 1990: 71) They do not have “SUBJECT(ENTITY) PLACE-LOCATIVE” construction representing existence as in Turkish.

- (7) *namd har saten uwei.*
 1SG.DAT black sugar no
 I don't have brown sugar.

Shiotani (1991: 53)

Note that the progressive for existence as in (5) differs from non-progressive as in (6) in terms of temporariness and mirativity. Example (5) implicates that “the stone” is not permanently there and represents that the speaker has just found “the stone”. Example (6) has two readings; one is that it asks where he lives (permanently) and the other is that it asks the listener, who knows the answer, where he is at that time.

In the negative context, the difference may merge into one form *uwei*. The progressive meaning is represented with the *uwei aa-ǰ+aa-wei* construction. The negative meaning with the past tense is represented with the *uwei aa-sen* construction. Unfortunately, we have not found the imperative or volitional formations (also in copular and in auxiliary use).

3.2 Copula *aa*

The copula *aa* is also used only in marked circumstances such as in past (8), progressive (9), or subordinate clauses. In an unmarked construction expressing “A is B,” no verbs are employed (10). In a negative sentence, *bišen* is used (11). In contrast, the existential verb *bai* in present Mongolian is often used as a copula with the unmarked form *bai-na*.

- (8) *udiš tenger yamer aa-sen?*
 yesterday sky how AA-PERF
 What was the weather yesterday?

Shiotani (1991: 65)

- (9) *šii ul dialle-ten=šii erin bas erd aa-ǰ+aa-wei.*
 2SG NEG to.be.late-PERF=2SG time yet early AA-SIM+AA-NPST
 You weren't late. It's early now.

- (10) *ter kuu ečig-min.*
 that person father-1SG
 That man is my father.

- (11) *šii minii meemee-mini bišin=šii.*
 2SG 1SG.GEN mother-1SG NEG=2SG
 You're not my mother.

The difference between (9) and (10) is the same as the counterpart of existential use (progressive (9) represents temporariness and non-progressive (10) permanent identity). Other constructions also follow it

3.3 Auxiliary *aa*

Auxiliary *aa* (or *+aa*), representing the progressive aspect, follows a main verb with a simultaneous converb suffix *-ǰ* similar to other auxiliary verbs.

In the two basic uses of *aa* as existential and copula, we cannot find the use of the unmarked verbal form of *aa-wei*. Only when this verb is used as an auxiliary verb, after the simultaneous converb form *-ǰ* expressing the progressive aspect meaning, is the unmarked form *aa-wei* used (12). However, the negative forms of *aa* in three uses, such as *ul aa* or *aa-gu uwei*, do not appear (13).

- (12) *bii tend iči-gu-d-mini ter wante-ǰ+aa-wei*
 1SG there to.go-VN-DAT-1SG that to.sleep-SIM+AA-NPST
 When I went there, he was sleeping. Shiotani (1991: 78)
- (13) *ter ugin keǰee eil-gu-ee-mel ul mede-ǰ+aa-wei*
 that girl when to.marry-VN-REFL-REFL NEG to.know-SIM+AA-NPST
 That girl doesn't know when she herself will marry. Shiotani (1991: 59)

They, of course, have the past formation with *-sen*, but we have not seen the examples with imperative or volitional.

According to the discription by Engkebatu (eds.) (1988), the *-ǰ+aa* formation is one element *-ǰaa*, as shown in the paradigm below.

	non-progressive	progressive
non-past	<i>-bei</i>	<i>-ǰaabei</i>
past	<i>-sen</i>	<i>-ǰaasen</i>

Table 1. Tense paradigm by Engkebatu (eds.) (1988: 308)

It is problematic that he deals with them as “tense” but the problem “*-ǰ+aa* or *-ǰaa*” deserves careful attention. This will be examined in the next section.

4. Auxiliary verb or suffix

Auxiliary construction with *aa*, *-ǰ+aa*, is possibly analyzed as suffix *-ǰaa*. On the basis of its accent, this construction sometimes, but not always, forms one word (In this paper, we express them as “compounds” by using “+”). We have to investigate its morphological features.

Section 3 demonstrates the following two points regarding the usage of this verb: first, *aa* does not have any negative forms in the usages (a)–(c); second, it does not appear as the “simple form” with the nonpast verb final form in (a) or (b) but does so in (c) as shown in table 2. Since we can use existential and copular verbs as verbal predicates, they have half-verbal properties; however, existential and copular constitute just a nominal predicates.

	(a) existential	(b) copula	(c) auxiliary	other verbs
negation	×	×	×	✓
simple form	×	×	✓	✓
past form	✓	✓	✓	✓
progressive form	✓	✓	✓	✓

Table 2. The usage of a verb *aa*

Auxiliary verbs do not always avoid negation. Some auxiliary verbs, like *šad* “can” and *uk* “to give,” accept negation.

- (14) *ter bas weer-ee warhl-aa ems-ǰ ul šad-n.*
 that yet self-REFL clothes-REFL to.wear-SIM NEG can-NPST2
 He hasn't been able to wear clothes by himself. Shiotani (1991: 83)

- (15) *ḡaa-ḡ ul uk-n=ṡie?*
 to.teach-SIM NEG to.give-NPST2=2SG.Q
 Could you tell me, please?

Shiotani (1991: 62)

On the other hand, negative word *ul* cannot be put between main verb and auxiliary verb *aa*. This means that the main verb and the auxiliary verb are strongly connected³.

The fact that auxiliary *aa* can take simple nonpast form and the comparatively strong connection between the two led us to the implication that the form *-ḡaa* may have a suffix-like feature⁴.

5. Cognate verbs with *aa* in other Mongolics

The verb *aa* is cognate with *a in Middle Mongol, and is attested in Modern Mongolian as in some fossil words. In this section, we see some examples in other Mongolics and try to summarize the cognates in other Mongolics.

In Central and Northern group, like Mongolian, Buryad, and Oirad, existential verb is replaced with **bayi* and the cognate verb with **bayi* is used in the three ways (16-18 Khalkha Mongolian).

- (16) *odoo urga-ḡ baj-gaa nogoo xaana=č baj-x=güj.*
 now to.grow.up-SIM to.be-IMPERFECT vegetable where=ever to.be-FUT=NEG
 We don't have well-grown vegetables anywhere. D. Namdag. *Cag törijn üjmeen*
- (17) *xar-san=čin' nodnin+ḡil manaj ang-ijn xural deer suu-san*
 to.see-PERF=2SG last.year our class-GEN meeting on to.sit-PERF
bagṡ mön baj-na.
 teacher just to.be-NPST
 I saw, he was just the teacher who sat for our class meeting last year.
 Š. Gaadamba. *Bagiin yawdal*
- (18) *talarxa-n temdegle-x-ijg xüs-č baj-na*
 to.thank-ASS to.celebrate-FUT-ACC to.hope-SIM to.be-NPST
 (They) are hoping to thank and to celebrate (it). L. Tüdew. *ḡargaliin duulal ḡaran bot'*

Their usage is similar to *aa* in Dagur, but existential verb **bayi* differs from *aa* in that they have the same negative form (16) as other verbs and that they are often used with nonpast verb final form in copular use (17).

The cognates with *aa* in Dagur are found in Monguor and maybe in Bonan, the South group of Mongolic, but they are more particle than verb (19-20).

- (19) Monguor
te jau-dz-a.
 that to.go-SIM-A
 He has left (I saw he left). Čenggeltei (eds.) (1988: 257)

³ But only the emphatic clitics =*l* can appear between them. In this case, we have to accept that they are two words (e.g. *gui-ḡ=l+aa-wei* {to.run-SIM=EMP+AA-NPST} 'run like crazy').

⁴ *-ḡaa* does not follow the vowel harmony but some suffixes such as volitional *-yaa* do not, either. We have no idea whether their close connection results from phonetic surroundings (e.g., the vowel-beginning feature) or semantic conditions (e.g., aspectual meaning, cf. *-ḡ tali* meaning perfective aspect).

(20) Bonan

ədzaŋ maɣei kuar naŋ-da əd-gu-a.
 3SG tomorrow city way-DAT to.go-FUT-A

He will go to city tomorrow.

Chen (eds.) (1986: 193)

In Middle Mongolian attested in *The Secret History of the Yuan Dynasty*, we find the verb *a*, meaning “to be,” “to live,” or “to sit,” apparently cognate with *aa* in Dagur. The verb *a* is used as an existential verb⁵ and auxiliary verb (21), but existential use differs from *aa* in Dagur as it can be used in the negative form⁶ (22).

(21) 說=着 坐 有
 客額=周 撒温 阿=木
 ke'e=jü sa'u=n a=mu.
 to.say=SIM to.sit=ASS A=NPST

(S/he) is sitting with saying something.

HuaYi YiYu, 3:02⁷

(22) 為長=着 不^○ 住麼 您
 阿中合刺=周 兀祿^兀 阿=[中]渾 塔
 aqala=ju ülü^ü: a=qun ta
 to.grow.up-SIM NEG A=VN 2SG

You have not grown up.

The Secret History of the Yuan Dynasty, 3: 49

Now we compare the verb *aa* in Dagur with cognates in other Mongolic languages (see table 3).

*a	Dagur	e.g. Mongolian	e.g. Monguor	Middle Mongol
form	aa	replaced with *bayi	-a	a
world class	verb	verb	gramaticalized	verb
ways of usage	existential, copula, auxiliary	existential, copula, auxiliary	existential, copula, auxiliary	existential, (maybe copula), auxiliary
negation	no	ok	no (not verb)	ok

Table 3. *a in Mongolic languages

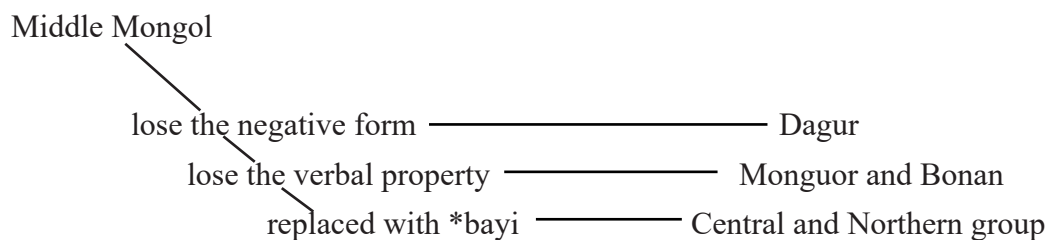
The verb *aa* in Dagur is closer with its counterpart of Middle Mongol because of its verbal property but it is different in negation. From this table 3, we can examine the historical changes as shown below (23).

⁵ It is hard to find copular use because of arbitrary semantic judgment.

⁶ The verb *a* in Middle Mongolian has some inflections, such as =ba, =duqai, =ju'u, =lu'a, and =mu. However, it does not mean that the verb *a* completes with the inflectional system; it rather lacks some inflection.

⁷ These examples are quoted from <http://hkuri.cneas.tohoku.ac.jp/> (last checked on 27/01/2017).

(23) Historical change from Middle Mongol to contempral Mongolics on *a⁸



6. Conclusion

This paper described the use of the verb *aa* in Dagur and the cognate *a in Mongolic languages. As same as typological view, existential verb is also used as copula and auxiliary and as is often the case, some of its usages differ from those of other verbs.

From a historical change viewpoint, *aa* in Dagur has lost its negative form. We can regard it as a process where *a lost its verbal property and became a particle as in Monguor and Bonan.

Abbreviations

-	suffix boundary	ASS	associative	PL	plural
=	clitics boundary	DAT	dative-locative	PROP	property
+	compound word boundary	EMP	emphasis	Q	question
1	1st person	FUT	future	REFL	reflexive
2	2nd person	GEN	genitive	SG	singular
3	3rd person	LIM	limitative	SIM	simultaneous
ACC	accusative	NEG	negative	VN	verbal noun
GA	genitive-accusative	NPST	non-past		
		PERF	perfect		

References

- Čenggeltei, eds. (1988). *mongɣur kele ba mongɣul kel / Tuzuyu he Mengguyu* [Monguor and Mongolian]. kökeqota: öbür mongɣul-un arad-un keblel-ün qoriy_a.
- Chen Naixiong, eds. (1986). *bou an kele ba mongɣul kel / Bao'anyu he Mengguyu* [Bonan and Mongolian]. kökeqota: öbür mongɣul-un arad-un keblel-ün qoriy_a.
- Engkebatu, eds. (1988). *dayur kele ba mongɣul kel / Dawo'eryu he Mengguyu* [Dagur and Mongolian]. kökeqota: öbür mongɣul-un arad-un keblel-ün qoriy_a.
- Shigeki, S. (1991). Daguurugo Hairaruhougenno kougoshiryōu -tekisutoto chuushaku- [Narrative Material in Hailar Dagur – text and note –]. *Nihon mongoru gakkai* 21, (1990). 47-95.

⁸ We need further consideration on *bū, another existential verb attested in Middle Mongol and preserved as some fossil words in Mongolic languages. *bū as a verb is not found neither in Dagur nor in the Central and Northern group. In Monguor and Bonan, the gramaticalized form *-ii / -i* is considered as the cognates with *bū.

A map of a region, possibly a country or a large administrative area, divided into several colored areas. The colors include orange, blue, yellow, light blue, and green. The map is tilted slightly to the right. The text "Part III: Pedagogy and Teaching" is overlaid on the map.

Part III: Pedagogy and Teaching

Language Factor in the Formation of National and Cultural Identity of Kazakhstan

Dina Andabayeva
Raushangul Avakova
al-Farabi Kazakh National University

Abstract

This article attempts to give an overview of the language situation and language planning in Kazakhstan. Statistical data is given and excursion to history of languages in Kazakhstan is done. The republic is home for more than 100 nationalities living in peace together. There is a great diversity of languages related to different language families. Particular emphasis is placed on the national - cultural component of Kazakh people namely the impact of the specificity of Kazakh language on ethnic identity. Language is one of the basic aspects of national identity. Recently in the Republic of Kazakhstan purposeful work on language development has been conducted. Optimal solution of language problems is a factor of interethnic relations harmonization, strengthening and consolidation of peoples and public consent. The evolution and maintenance of languages in Kazakhstan take an important place in state policy. The problem of the state language as part of national (civil) identification play a huge role in successful integration process of Kazakh society. And quite rightly assume that one of the foundations of a new civic identity is knowing Kazakh language by all citizens of Kazakhstan. The article is an analysis of language situation in Kazakhstan in connection with historical and political peculiarities.

Keywords: ethnolinguistics, Kazakhstan, language personality, language planning, mentality.

1. Introduction

In the Republic of Kazakhstan, which is a multi-ethnic state, the current problem is the formation of multicultural person, who owns several languages. President Nursultan Nazarbayev in his 'Message to the Nation' set the task of developing multilingual education society, which can serve as a reliable basis for building a competitive state, which is planned to be one of the 50 most competitive countries of the world.

Polylingual person must have national sense, civic consciousness, patriotism and love for homeland on the basis of knowledge of traditions of native culture and tolerance for other cultures. Therefore, one of the important tasks of the school is to build a high level practical knowledge of different languages. The Republic of Kazakhstan is multinational state, which means interaction of cultural values of different nations. The state language of the country is Kazakh; Russian is official language and can be used in paperwork at the same level as Kazakh.

The concept of language policy in Kazakhstan determines its main difficulty in creating an optimal language area of the state, requiring a clear definition of the functional relation of languages in which the state language should take its rightful place. Kazakh President Nursultan Nazarbayev believes that national diaspora can act in 'arrayed common Kazakh cultural reality not as broken crops of their ethnic and continent cultures, but become authorized representatives of Kazakhstan, one of the channels of relationship between our country and national and cultural systems of other states' (Nazarbayev, 1999).

National communities living in the newly independent state - the Republic of Kazakhstan, have an important mission to act as agents of cultural dialogue between Kazakhstan and other countries. At the same

time ethno-linguistic communities in Kazakhstan show its political, linguistic, ethnic subjectivity through their cultural development and cultural activities, including exit to the level of international relations.

It is noted that despite the political and other influence on the process and results of the census realization as well as objective difficulties in the sociolinguistic identification and inventory of languages, the list of 126 languages spoken in Kazakhstan which in accordance with the status and spread in Kazakhstan and in the main country of residence of reference peoples were divided into the following groups : (1) State Kazakh language, (2) Russian (3) Exogenous and endogenous languages of diaspora (4) Immigrant languages (Suleimenova, Shaimerdenova, Akanova, 2007)

It is quite obvious that with such a huge number of ethnic groups living in the territory of Kazakhstan it is necessary to pay special attention to the issues of integration and interaction between them.

In this regard, Kazakhstan has created a unique in its structure body - the Assembly of Peoples of Kazakhstan as an advisory body as well as scientific advisory council to the Assembly which conducts a huge work in the spheres of all processes associated with ethno-linguistic, historical and cultural components of the current political situation of Kazakhstan.

How can we characterize the ethnolinguistic situation of modern Kazakhstan and what are the prerequisites for further development of languages and cultures in our country?

First of all, scientists have noted a steady Turkic substrate which was formed on the territory of Kazakhstan in the form of the titular ethnic group language - Kazakh language which is official language expanding its functions and is an important component of language situation and language development in Kazakhstan (Constitution of the Republic of Kazakhstan , Art. 7, paragraph 2).

2. Research Methods

The methodological basis of the study is introduced in methods used in the modern knowledge of social and political processes and phenomena:

- historical method, which allows to consider the problem in retrospect, to make a comparative review of the language situation in Kazakhstan from past to present, to determine the dynamics of the development of the state language;

- structural and functional methods help to analyze the language policy as a part of the government and one of the functions of the state activity;

- method of application analysis is a tool that was used to assess trends and forecasting prospects for the development of language policy in the Republic of Kazakhstan.

- method of questionnaire survey served the purpose to collect sociolinguistic material

- method of quantitative analysis was used for data processing

- method of analysis of written sources: official documents, materials of population census, statistical compilations and mass media. Of great importance was the use of

- descriptive method, including methods of internal and external changes in the interpretation of the facts of the language situation.

3. Discussion

Currently there are 126 languages in Kazakhstan. Kazakh is autochthonous language, other languages are of dispersed living non-indigenous ethnic groups of diaspora: Russian, Uzbek, Kyrgyz, Tajik, Bashkir, Tatar,

Azeri, Ukrainian, Polish, Czech, German, Bulgarian, Greek, Korean, etc.

For a long period of time, Kazakhstan was a part of the Russian Empire and later the Soviet Union. That time fundamental and the de facto government language throughout the country was Russian until the mid of 1980s abruptly (in 1939, 1957, 1969, 1983) the paperwork in Kazakh language was folding in rural areas (originally paperwork in cities was in Russian) and there was active closure of Kazakh language schools.

The adopted Law of Languages on September 22, 1989 in the Kazakh Soviet Socialistic Republic (KazSSR) the concept of 'state language' was introduced and Kazakh was given the status of state language, for the Russian language de jure status as the 'language of interethnic communication' was secured. Later with amendments to the Constitution of the Republic of Kazakhstan was determined that 'along with the state language Russian is official language' (The Law 'On languages in the Republic of Kazakhstan', 1997).

On May 15, 2002 at the presentation of the public association 'Russian Party of Kazakhstan' (RPK) it was announced that the primary task of the RPK - is to call on all Russians living in Kazakhstan to study the state language. As soon as the Russian Party was opened more than 3 thousand people became its members.

Ministry of Agriculture of the Republic of Kazakhstan fully translated all documentation into the state language according to the order of Minister of Agriculture of the Republic of Kazakhstan dated July 5, 2004, # 348.

In 2004, according to the assessment of D. O'Biken, Kazakh government has decided that Kazakh language will play a crucial role in the process of so called 'kazahization' of society (O'biken, 2005).

Thanks to the work of the Assembly of Peoples of Kazakhstan, languages and cultures having no official status in Kazakhstan are highly supported (or at least do not hinder their development).

On September 12, 2005 speaking on the second Civic Forum, the President of the Republic of Kazakhstan Nursultan Nazarbayev said: 'We must work together to support the languages and cultures of all peoples of Kazakhstan. No one will be discriminated in their rights to use their native language and culture' (www.akorda.kz).

By 2006, according to the voiced data at the round table 'The Future of Kazakhstan and the state language', paperwork was officially translated into Kazakh language in five areas (Atyrau, Zhambyl, Kyzylorda, Mangistau and South Kazakhstan regions). Nevertheless, even in these areas the 'kazahization' of official documents faced several challenges. Thus, in the Atyrau region only 51 % of the official documents were carried out in Kazakh language; in Jambul - 50%, in Kyzylorda - 49 %. In 2006, 500 million tenge from the republican budget was allocated to the introduction of Kazakh language.

In 2007 the number of citizens of Kazakhstan with high knowledge of state language has reached about 70 percent. According to the Minister of Culture and Information Ermukhamet Ertysbayev 10 years ago when the Law 'On languages in the Republic of Kazakhstan' was adopted this number did not exceed 40 percent.

At the same year cultural project 'Trinity of languages' - Kazakh, Russian and English in Kazakhstan was adopted on the state level. However, this idea had both supporters and opponents.

On August 20, 2007 at the XIII session of the Assembly of Peoples of Kazakhstan, the Head of the State Nursultan Nazarbayev stressed:

'We must make every effort for the further development of Kazakh language which is the main factor uniting all the peoples of Kazakhstan. At the same time it is necessary to create favorable conditions for persons of all nationalities living in the country to speak freely and being able to be taught their native language and develop it'.

In early September 2009 it was reported that by 2010 official documentation in Kazakhstan would be translated into official language by 60% rather than 70%, as planned. It was due to the fact that at the beginning of September the Government of the Republic decided to change the Strategic Plan of the Ministry of Culture

and Information of the Republic of Kazakhstan for 2009-2011.

Minister of Education and Science of the Republic of Kazakhstan Zh.Tuimebayev in his interview on January 29, 2010 on the issue of translation the official documentation into Kazakh language from 2010 said that ‘All official documentation is maintained in national Kazakh and sub-national Russian languages. No displacement is expected. This was also recently announced by our President Nursultan Nazarbayev’. He also confirmed that Russian language will retain all functions of the language of interethnic communication: ‘I therefore emphasize that the rumors and speculation about a displacement of the Russian language is incorrect’. Russian language - is our heritage, and our president has repeatedly talked about this. Kazakhs learned the world culture through the Russian language. There are about 30 percent of Russian schools in Kazakhstan, i.e. those where all subjects are in Russian.

In 2012 under the amendments to the Law ‘On Culture’ all films imported to the territory of Kazakhstan must be dubbed into Kazakh language.

On August 4, 2011 the Ministry of Culture of the Republic of Kazakhstan has drafted a bill which stated that: ‘Applications (complaints) to public authorities and public bodies responses to the treatment of individuals and entities and other documents maintaining accounting, statistical, financial, technical and other documents in the system of government of the Republic of Kazakhstan as well as all documents emanating from government must necessarily be in Kazakh language. And for doing accounting, statistical, financial, technical and other documentation in organizations of Kazakhstan a reservation was made, this information is provided in Kazakh language and only “if necessary” - in Russian. Names of legal persons and objects from the date of adoption of the law will be given only in Kazakh language. In education, the number of Kazakh groups and classes must be at least 50%’.

On September 6, 2011 the public persons demanded from the President of Kazakhstan to exclude from the Constitution the item about the usage of Russian language. This demand was signed by: the co-chair of the party ‘Azat’ Bulat Abilov, the chairman of the party ‘Ak Jol’ Azat Peruashev, Soviet cosmonaut Tokhtar Aubakirov, Honoured Artist of the Soviet Union Bibigul Tulegenova, the chairman of the Green Party ‘Rukhaniyat’ Serikjan Mambetalin, as well as writers, journalists, and many others. In response to this movement on September 7, 2011 the advisor of the President of Kazakhstan on political issues Ermukhamet Ertysbayev answered to the letter of public persons with the statement that the ‘vast majority of Kazakh voters oppose such a position’ and added that ‘Nobody and never will be able to dislodge Russian language’.

On December 14, 2012 ‘Kazakhstan by the year of 2025 intends to translate the alphabet from Cyrillic to Latin’, – said the President Nursultan Nazarbayev, giving his speech at a solemn meeting dedicated to the 21st anniversary of independence. ‘It is necessary to start a preparatory work on the transition of Kazakh alphabet to the Latin by the 2025’ - said Nazarbayev. According to the president Latin alphabet will turn Kazakh into ‘the language of modern media’. By that time, in 2025 Kazakh language must become ubiquitous and dominant in all spheres of national life.

Every year in September the Festival of languages is held on the territory of Kazakhstan. This festival is important event for all the peoples living in Kazakhstan as it contributes to the strengthening of friendship and unity of the people.

Currently Kazakh scientists are carrying out a great job on learning Kazakh, Russian, Uzbek, Uighur and many other languages .

The most important issues that excite philologists are:

Kazakh language as the state language in the context of the linguistic situation of the Republic of Kazakhstan.

Russian as mother language, as second language and as foreign language.

Languages of ethnic groups, such as Ingush, German, Korean, Uighur and many other languages of Kazakhstan.

Communicative space of languages and cultures.

Linguistic identity and tolerance.

Linguistic regulation in Kazakhstan is held in the context of multiethnic and multicultural society and aimed at optimizing of interethnic relations and formation of ethno-linguistic tolerance, development of interethnic integration and strengthening the unity and integrity of society. 'The language policy is designed to create a harmonious sociolinguistic space in the country, to become a factor of unity of the people. It should be aimed at broadening and strengthening social and communicative functions of the state language; Russian language dialog function preservation; development of the languages of other ethnic groups' (Doctrine of National Unity, 2009).

In modern conditions of statehood, prospects are directly dependent on the state's ability to maintain and protect the national system of values. The protection of national values is the main purpose of measures to ensure national security. Loss of state identity, lack of unified database of values is a threat to national security. As K. Colin noted: 'One of the important factors in ensuring national security is the state language' (Colin, 1996).

Being a critical threat to society information and psychological impact on the consciousness of a human and people in general is carried out mainly through the language environment and through the language itself.

According to L. Puhovich, the question of state language as a part of national identity is a definite threat to security and stability of the country. At the same time the successful solution of the language problem can give a powerful impetus to further development of democracy and create the preconditions for entering the country's fifty most competitive countries of the world (Puhovich, 2007).

That is why the active defense of the language environment is an important aspect of national security. Language is one of the basic aspects of national identity.

Recently a determined purposeful work on language development is being done in the Republic of Kazakhstan. Optimal solution of language problems is a factor of harmonization of interethnic relations, strengthening and consolidation of peoples and public consent. Development of languages - one of the important directions of the state policy in the Republic of Kazakhstan (Functioning and development of languages state program of the Republic of Kazakhstan for 1998-2000).

As B. Abdygaliev said: 'One of the foundations of new national identity is knowledge of Kazakh language by all citizens of Kazakhstan. The key to stability and prosperity is linguistic unity of Kazakhstani when Kazakh and Russian languages are symbols of mutual understanding and respect. Currently Kazakh as the state language has high expectations. Thus, priority is given to the expansion of social, communicative and consolidating functions of the state language' (Makulbekov, 2010).

As explained by Puhovich L., further use of Kazakh language by representatives of all ethnic groups of Kazakhstan will not be a policy of assimilation as some Russians fear, it will only create the preconditions for the formation of national identity in the country. Experience in a number of European and Asian countries shows that fluency in two or three languages becomes not the exception but the rule.

The answer to this question we will search in the results of a poll conducted in 2010 by the Centre for the study of interethnic relations of the Republic of Kazakhstan. Sociological survey covered all regions of Kazakhstan. The total number of respondents was 1,200.

According to the survey it appears that 41.6 % of Kazakhstanis think in Kazakh language, while 32.1 % of respondents think in Russian; the answer "We can think both in Kazakh and Russian languages" gave 15.5 % of respondents. This question cannot be overlooked without the latest data of the Agency for Statistics of the Republic of Kazakhstan that the major share of the population in the country over 64 % are Kazakhs, the second

highest number are Russians (24%), 2.91 % are Uzbeks, 1.95% are Ukrainians, 1.42 % are Uighurs, 1.24 % are Tatars, 1.10 % are Germans, 4.51 % are other ethnic groups (Demographic Yearbook of Kazakhstan Statistical compilation, 2011).

To the question ‘Do you think that the language environment of the state language has been expanded today?’ the majority of respondents (30.5%) believe that it has been expanded throughout, especially after the release of the State Language Law. 19.9 % of the population claim that the language environment of state language expanded only at official level, in official documentation, while 18.5 % believe that educational institutions begin to allocate more training hours to state language study (see Table 1).

Answers	%
Yes, it has been expanded throughout, especially after the release of State Language Law	30,5%
Yes, expanded only at official level, in official documentation	19,9%
Yes, educational institutions begin to allocate more training hours to state language study	18,5%
Yes, media (TV programs, periodicals) give more airtime for programs in official language	17,8%
No, hardly expanded, changes are of ‘cosmetic’ nature	12,5%

Table 1. Do you think that the language environment of state language has been expanded today?

According to the results, 17.8 % of respondents believe that ‘Media (TV program, periodicals) give more airtime for programs in state language’. This circumstance can be proved by the fact that the biggest state television channel ‘Kazakhstan’ by decision of the president of the Republic of Kazakhstan Nursultan Nazarbayev on September 1 switched to the broadcasting in Kazakh language.

‘When hiring to a job, knowledge of state language was required, which created some difficulties for me’ - believe 20.4 % of respondents; ‘Difficulties were connected with filling some documents which required knowledge of the language’ - 12.4%, ‘Did not experience any difficulties because of Kazakh language became official language’ indicated the greatest number of respondents (45.9 %).

Today 29 national and 16 local executive bodies transferred paperwork to state language. As part of the development of infrastructure for teaching the state language a unified network of the state language teaching organization has been created for adult population including civil servants. Currently, there are 93 state language learning Centers. State language teaching courses are working in all central and local government bodies.

Recently ‘The Development and Functioning of Languages’ state program for 2011-2020 was adopted. The aim of the Program is harmonious language policy that provides full operation of the state language as the most important factor in strengthening national unity while preserving the languages of all ethnic groups living in Kazakhstan. Kazakh language along with Russian and languages of other ethnic groups continue to operate and develop. Conditions are being created and every opportunity is being used to meet national and linguistic queries of citizens. In this regard a wide network of Sunday schools was organized.

Results on the level of Kazakh language proficiency in Kazakhstan are presented in Figure 1.

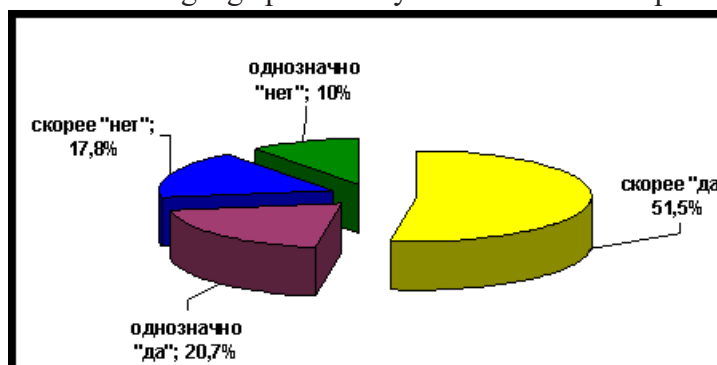


Figure 1. Are you satisfied with your knowledge of Kazakh language?

Blue pie-chart – ‘probably not’

Purple pie-chart – ‘definitely yes’

Green pie-chart – ‘definitely no’

Yellow pie-chart – ‘probably yes’

As we see from the Figure 1, 51.5 % of Kazakhstanis are satisfied with knowledge of Kazakh language, the answer ‘definitely yes’ chose 20.7% of respondents, not contentment knowledge of Kazakh language is expressed in response ‘definitely no’ (10%); 17,8 % of respondents are not happy with their knowledge of state language.

Since gaining long-awaited independence it is still debated about what Kazakh as the state language content should be. The position of Kazakhstan aborigines’ language might be dramatically improved as it became not only de jure but also de facto official language of the republic. Citizens were asked ‘In your opinion what kind of content (primarily) is needed to invest in the concept of Kazakh language?’ According to the survey 45.1 % of Kazakhstanis believe that primarily Kazakh language - is a state language, 24.1 % of respondents said that Kazakh language - is a factor of national identity , the most important element of the cultural values of Kazakh ethnic group, whereas 16.4% perceived the Kazakh language as a communication tool, ‘Kazakh language (as well as languages of other ethnic groups) - is soul of the people’ believe 13.7 % of citizens of Kazakhstan.

4. Conclusion

Thus, the problem of state language as part of national (civil) identification play a huge role in successful integration process of Kazakh society. And quite rightly to assume that one of the foundations of new civic identity is knowing Kazakh language by all citizens of Kazakhstan that is shown in the poll results. The key to stability and prosperity is linguistic unity of Kazakhstanian when Kazakh and Russian languages are symbols of mutual understanding and respect (Abdygaliev, 2007).

The question of knowledge of Kazakh language is always associated with the formation of Kazakhstan patriotism, since knowledge of Kazakh language is primarily knowledge of state language of the country. Today all the possibilities for learning Kazakh language are being created. In state institutions and enterprises courses are opened, scientific and methodological materials are published and distributed. By the order of the President, employees of government agencies who are non-Kazakh nationalities but owning state language are financially encouraged. Many of our fellow citizens are aware of their patriotic duty - to learn Kazakh language. In addition, it is understood that state language is a factor of personal competitiveness and career advancement in any field (Savin, 2009).

References

- Abdygaliev B. *Kazakh mission // Kazakhstan’s Pravda*, July 25, 2007.
- Colin, K. (1996). *On the role of the Russian language in the preservation of a unified information space of Russia*. Information Society, vol. 6.
- Constitution of the Republic of Kazakhstan*, Art. 7, paragraph 2.
- Demographic Yearbook of Kazakhstan Statistical compilation / 2011-* 592 p . Edited by Smailova A.
- Doctrines of National Unity. Astana. (2009) (Draft).
- Functioning and development of languages state program of the Republic of Kazakhstan for 1998-2000*, N 4106.
- Makulbekov, M. (2010). *The Ministry of Culture work for the further strengthening of the statehood // News agency Kazinform*, Astana.

- Nazarbayev, N. (1999). *In the flow of history*. Atamura, Almaty.
- O'biken, D. (2005). *Kazahization and language policy in post-Soviet Kazakhstan // Political science and government in the Russian Federation and the Newly Independent States*. Ural Branch of RAS, Ekaterinburg.
- Puhovich, L. (2007). State language and national identity as factors of security and stability in Kazakhstan // Security : international, regional, national (system approach). *Proceedings of the international scientific-practical conference dedicated to the memory of Doctor of Political Sciences, Professor MS Masha. Dyke –Press, Almaty.*
- Savin J. *Ethics as a factor of everyday life in the rural areas of South Kazakhstan* NZ, № 4, 2009
Statement of the President of the Republic of Kazakhstan Nursultan Nazarbayev at second Civic Forum - akorda.kz.
- Suleimenova, E., Shaimerdenova, N., Akanova, D. (2007). *Languages of the Peoples of Kazakhstan: Astana sociolinguistic Directory*, Astana.
- The Law of the Republic of Kazakhstan* 'On languages in the Republic of Kazakhstan' (from July 11, 1997 # 151-1, as amended by Law dated December 20, 2004 # 13 -III). // Kazakhstan's Pravda , July 16, 1997.

Project-Based Language Learning (PBL) Instruction in an Online Language Learning Environment: A case study in Pashto

Rahman Arman
Amber Kennedy Kent
Öner Özçelik
*Center for Languages of the Central Asian Region (CeLCAR)
Indiana University*

Abstract

Throughout the past century, project based learning (PBL) has proven itself to be a formidable instructional mode for producing engaging learning environments across many disciplines by taking learners to design and construct actual solutions to real-life problems. However, there has thus far been limited research on the effectiveness of PBL in the field of language learning teaching, with is particularly true for the sub-field of lesson commonly taught languages (LCTLs). This paper describes a comparison study where we followed the Gold Standards of PBL Design to design two parallel courses teaching the same defined learning outcomes and PBL assignment but using two different methods of delivery: face to face and online. In doing so, we seek to examine the benefits of PBL in both the traditional language classroom and an online language learning environment.

Keywords: project based language learning, online language learning, LCTLs

1. Introduction

Project-based learning (PBL) has proven itself a formidable instructional method for producing engaging learning environments across many disciplines by tasking learners to design and construct actual solutions to real life problems. Recently, there has been an increased use of PBL in the language classroom, commonly referred to as *project-based language learning* (PBL) (Mikulec and Miller, 2011); however, there is still limited research on the effectiveness of PBL in the sub-field of less commonly taught languages (LCTLs) and non-traditional instruction, such as online or distance language learning environments. As such, this study asks: *Does using PBL in an online Pashto language course promote oral fluency through increased learner engagement and autonomy?*

2. Context

The project was implemented with 15 learners across five semesters at Indiana University as shown in Table 1.

Semester	Sum 16	Fall 16	Spr 17*	Sum 17	Total
Undergraduate	1	3	3	3	7
Graduate	2	4	4	2	8

**The same learners enrolled in Introductory I Pashto in Fall 2016 continued to Introductory II Pashto in Spring 2017.*

Table 1: Summary of student enrollment

3. Methods

We used a mixed methods approach, analyzing both the *quantitative* results of formative assessments using classroom assignments and a summative assessment via an Interagency Language Roundtable oral proficiency interview (ILR OPI) to analyze the learner’s proficiency levels and fluency, as well as *qualitative* results of a post-course survey in which learners rate the course according to the Gold Standards of PBLL as well as indicate their perceived levels of interest, engagement, motivation, and performance.

We chose to use the ILR proficiency standards for summative assessment for this project because we have found the majority of learners studying Pashto have a goal of working for agencies that use these standards for evaluating proficiency (both government and NGOs). Figure 1 illustrates how the ILR proficiency standards relate to the commonly accepted *Standards for foreign language learning* as developed by the American Council on the Teaching of Foreign Languages (ACTFL), which is generally the industry standard for most language education training in higher education in the United States, and Table 2 below summarizes the proficiency levels we aimed for as part of this PBLL case study.

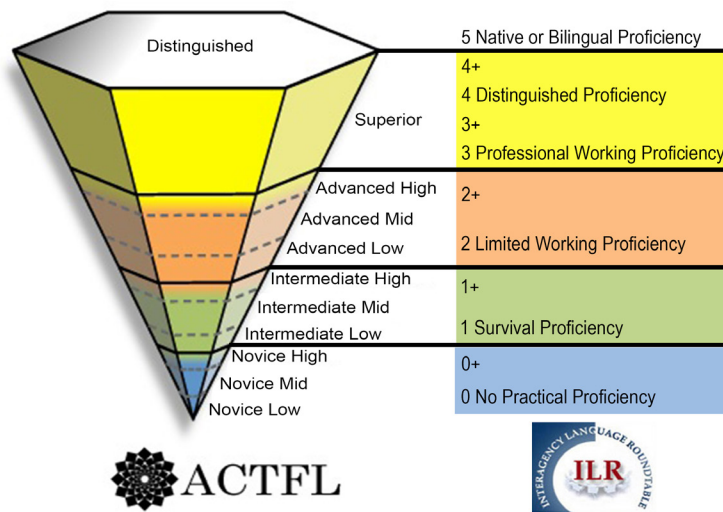


Figure 1: Comparison of ACTFL and IRL Proficiency Standards

ACTFL Rating	ILR Rating	Description
Intermediate <i>Low-Mid-High</i>	1/1+ <i>Survival Proficiency</i>	Able to create language on familiar topics related to their daily life. Can recombine learned material to express personal meaning. Can ask simple questions and handle straightforward situations. Produce sentence-level language, ranging from discrete sentences to strings of sentences, typically in present time.
Advanced <i>Low-Mid-High</i>	2/2+ <i>Limited Working Proficiency</i>	Can communicate autobiographical information, as well as discuss topics of community, national, or international interest through narration and description in the major times frames. Can also deal with unexpected complications.

Table 2: Proficiency levels aimed for at the complete of the Pashto courses

4. Gold Standards of PBL Design

Before designing a PBL assignment for language learning we reviewed work done by the **Gold Standards of PBL** as developed by the Buck Institute for Education (BIE), considered the **World-Readiness Standards for Learning Languages** (sometimes referred to as “the five Cs”) as defined by ACTFL (Figure 2), and incorporated lessons and discussion led by the National Foreign Language Resource Center (NFLRC) as part of their *2016 Online Symposium: Project-Based Language Learning: Inspiring Teachers, Transforming Learning* to create a list of Gold Standards of PBL Design as show in Table 3. These Gold Standards of PBL are what we followed to create and implement our own PBL lesson.



Figure 2: World-Readiness Standards for Learning Languages

GOLD STANDARDS OF PBL DESIGN	
1. Challenging Problem or Question	Present a meaningful and engaging question about a real-world problem used to increase proficiency in the L2 and engage with the target culture.
2. Sustained Inquiry	Facilitate sustained inquiry in the L2 over an extended period of time that delves into multiple layers of a topic, engages in meaningful communications, and spurs intercultural connections.
3. Authenticity	Integrate authentic texts, while engaging in authentic communication and promoting authentic evaluation while solving a real-world problem and making a real-world impact.
4. Student Voice & Choice	Give student choices in the design, creation, and presentation of the public project and allow for learner autonomy by leading students to define their own language learning goals.
5. Reflection	Promote and provide significant opportunities for learners to reflect on language and cultural knowledge and raise awareness of their own communicative and culture competence in the process.
6. Critique & Revision	Use scaffolding and modeling plus constructive feedback on language use (instructor) as well as project content (instructor, peers, and external experts) guided by clearly designed assessment rubrics.
7. Public Product	Create a public product designed for consumption by speakers of the L2 which demonstrates both target language and cultural competencies.

Table 3: Gold Standards of PBL Design

5. PBL Project Description

5.1 Problem/Challenge

As shown in Table 3, a main tenant of PBL is to begin with a question about an intriguing conceptual issue, concerning social problem, or school or community challenge, etc. that is relevant to your learner audience that they will work together to solve using sustained inquiry over the course of the semester. For our problem and/or challenge, we chose: *How do we help graduating and newly graduated Pashto learners locate employment both in the public and private sectors home and abroad utilizing their skills as Pashto speakers?*

5.2 Purpose

Next, it is important to explain to the learners why it is important to work on the assigned problem, including how participating in the project will help the class participants, school, community, or the world. We provided the following statement of purpose for our learners:

Pashto learners often find themselves challenged to secure relevant employment after graduation. The opportunities are plentiful, but the graduate often lack the resources and/or connections to locate them.

Learners will work together to help solve this problem of ***Finding Job Opportunities for Pashto Speakers***, by creating an exhaustive resource for soon to be and newly graduated students looking for employment utilizing their language proficiency skills.

5.3 For Public Product

An essential component of a well-developed PBL is a final product consumable by a predefined public audience of L1 or L2 speakers and others. This product should be developed so that it is actually demonstrated and/or shared with the intended audience within the timeframe of the course (i.e. a public produce that is developed but then never shared cannot be evaluated for its effectiveness). We required the following public product for the learners to develop cooperatively as a team.

The final project will be a working online resource that should include at a minimum:

- list of potential employers (including descriptions and contact information)
- sample resumes/CVs and cover letters along with tips and tricks on writing these
- resources on the most widely accept proficiency exams
- interview preparation videos.

Other areas of interest can be added by students as they research and develop the online tool.

5.4 Public Audience

Finally, it is essential to define for the learners who is the audience who must benefit from the information gleaned during the process of the project and the use of the final product, so that they can accurately develop a solution that will meet the needs of the intended audience. We defined the following required audience for our learners to consider while solving their problem and developing their public product.

The audience for the final project is:

1. Soon to graduate, graduating, and newly graduating Pashto learners (employment seeking tool)
2. Language instructors (marketing tool to engage interested students and a retention tool encouraging learners to continue their studies and increase their proficiency)
3. Potential employers (recruiting tool for finding new graduates with the desired language proficiency skills)

6. Results

Typically, at the end of one semester of Pashto (Intro I), we expect learners to score a 1 on the ILR OPI exam and learners who complete two semesters (Intro II) to score a 2 on the IRL OPI. At the completion of our PBLL courses, 100% of the learners who completed only Intro I performed higher than their typical non-PBLL peers and 43% of the learners who completed both Intro I and Intro II performed higher than their non-PBLL peers as shown in Tables 4 and 5 below.

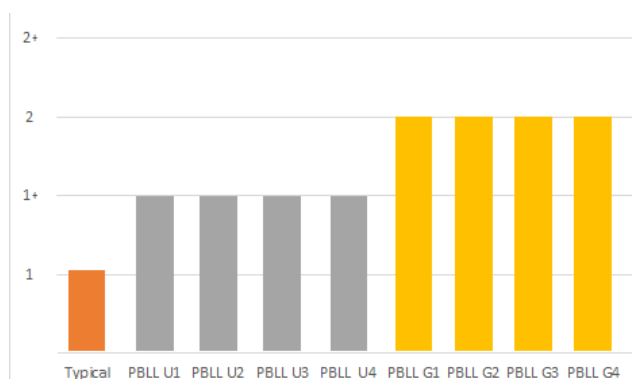


Table 4: Introductory I only ILR scores

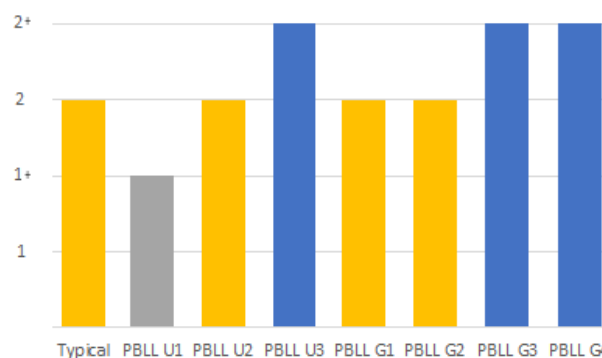


Table 5: Introductory I and Introductory II ILR scores

Additionally, survey respondents unanimously reported they felt the project was meaningful to their language learning experience and personal objectives enjoyed participating in the project felt their language learning experience and proficiency level benefited from the inclusion of this project would recommend including PBLL in future Pashto language courses.

7. Reflections

Overall, our findings were consistent with what we expected based on our hypothesis about incorporating PBLL. However, we were surprised that based on our small sample, it was the Intro I learners who seemed to benefit the most from PBLL, with *all* the Intro I-only learners performing higher than normal on the ILR OPI. The undergraduate participants scored 1+, instead of the typical 1 and the graduate participants scored a 2 at the end of Intro I, the score usually expected of learners at the end of the Intro II course!

Our second most significant finding was the self-reported increase of interest and motivation via the post-course questionnaires. We expected learners to be highly interested and motivated by the content matter, but we were especially pleased to find that learners unanimously reported feeling more motivated and interested in this project, recommending the continued use in future courses. As a result of this study, we will be continuing the use of PBLL in our future LCTL online course development.

References

- American Council on the Teaching of Foreign Languages (ACTFL). (1996). *Standards for foreign language learning: Preparing for the 21st century*. Yonkers, NY: ACTFL.
- American Council on the Teaching of Foreign Languages (ACTFL). *World-Readiness Standards for Learning Languages*. Retrieved from <http://aappl.actfl.org/> (accessed October 2, 2016).

- Buck Institute for Education. *Gold Standard PBL: Essential Project Design Elements*. Retrieved from <https://www.pblworks.org/what-is-pbl/gold-standard-project-design> (accessed October 2, 2016).
- Interagency Language Roundtable. *ILR Proficiency Standards*: Retrieved from <http://www.govtilr.org/Skills/ILRscale1.htm> (accessed October 2, 2016).
- Mikulec, E. and Miller, P. (2011). Using Project-Based Instruction to Meet Foreign Language Standards. *The Clearing House*, 84; 81-86. UK: Taylor & Francis Group.

Noun Phrases in Lhasa Tibetan

Yunchuan Chen

L. Julie Jiang

University of Hawai‘i at Mānoa

Abstract

This paper provides a detailed description of the syntactic distribution and semantic interpretations of three types of noun phrases, bare nouns, number constructions and demonstrative phrases, in Lhasa Tibetan, one of the Tibetan dialects spoken in an area of Central Tibet surrounding the city of Lhasa, China. We first show that bare nouns in Lhasa Tibetan share the same properties as those in classifier languages such as Mandarin. Second, we show that unlike Mandarin, numerals combine directly with bare nouns in Lhasa Tibetan in the same way as those in number marking languages like English and Hindi. However, unlike English or Hindi, nouns in Lhasa Tibetan are not marked with number morphology and can be interpreted as singular or plural. At same time, we will see that numeral constructions in Lhasa Tibetan exhibit the same six properties as those in number marking languages and classifier languages as observed in Jiang (2012). Third, we provide two diagnostic tests from Löbner’s (1985) and Alexiadou et al. (2007) to argue that the so-called definite determiner *de/di* in Lhasa Tibetan are indeed demonstratives, which can appear in two different positions in the numeral constructions.

We explain the particular properties of the three types of noun phrases in Lhasa Tibetan with the approach to bare nominals by Chierchia (1998) and Dayal (2004). The account of noun phrases in Lhasa Tibetan bears on current discussions about the nature of argument formation and contributes to the discussion about language variation and argument formation in general.

Keywords: Tibetan, nominal arguments, argument formation, bare nouns, numerals

1. Introduction

We begin this paper by presenting properties of three types of noun phrases, bare nouns, numeral constructions, and demonstrative phrases, in Lhasa Tibetan, one of the Tibetan dialects spoken in an area of Central Tibet surrounding the city of Lhasa, China.¹ It will be shown that Lhasa Tibetan is neither a number marking language, that is, a language with morphological exponents of grammatical number, such as English and Hindi, nor a classifier language, i.e. a language that has an obligatory system of classifiers, like Mandarin. At same time, it will be shown that Lhasa Tibetan exhibits the characteristics of both types of languages. The analysis of the structure and semantics of the three types of noun phrases presented in Section 2 will be taken up in Sections 3. In Section 4, the paper ends with a summarizing conclusion.

2. Properties of Noun Phrases in Lhasa Tibetan

In this section, we will present properties of three types of noun phrases, bare nouns, numeral constructions and demonstrative phrases, in Lhasa Tibetan.

¹ Tibetan belongs to the Bodic branch of Tibeto-Burman languages in the Sino-Tibetan language family (e.g. DeLancey, 2003, Shafer, 1955). The most comprehensive current classification of dialects of Tibetan is that of Nishi (1986), who distinguished six major groups: Central or U-Tsang (Lhasa, Shigatse, Sherpa, Kagate, etc.), Western Archaic (Balti, Ladakhi, Purik), Western Innovative (Lahul, Spiti), Southern (dialects of Sikkim and Bhutan), Khams, and Amdo (cf. DeLancey 2003, p.270). Lhasa Tibetan belongs to the Central group and is the best-known contemporary Tibetan dialect, with a SOV word order (DeLancey, 2003, p. 270-283; Denwood, 1999). Most data from Lhasa Tibetan in this paper was collected by our consultant work with several native Lhasa Tibetan speakers. Data from the literature will be duly noted.

- (4) ngas sgrasdudphruchas gcig dang cong gcig nyos payin. sgrasdudphrulchas thabstshang
 I-ERG radio one and clock one buy PST radio kitchen
 nang la bzhag yod. cong mgronkhang nang la bzhag yod.
 inside at put PERF clock lobby inside at put PERF

‘I bought a radio and a clock. The radio is in the kitchen and the clock is in the lobby.’

Finally, the definite reading of bare nouns in Lhasa Tibetan is supported by the example in (5) in which the bare noun *zhimis* ‘cat’ can receive a definite interpretation when the referents are visible and uniquely identifiable for both the speaker and the hearer:

- (5) Context: both the speaker and the hearer are looking at the scenario in which a cat or a group of cats are running to catch fish in the river.

zhimis gtsangpo-i nang-la nya zabrtsis dug.
 cat-ERG river-GEN inside-TO fish catch would

‘The cat/cats will catch fish in the river.’

We saw that bare nouns in Lhasa Tibetan share the same properties as those in classifier languages such as Mandarin; in the following subsection we examine numeral constructions in Lhasa Tibetan.

2.2 Numeral Constructions in Lhasa Tibetan

In Lhasa Tibetan, bare nouns combine with numerals directly:

- (6) deb gcig/ gnyis
 book one/two
 ‘one book/two books’

The above examples distinguish Lhasa Tibetan from classifier languages like Mandarin which obligatorily require a classifier to connect bare nouns with numerals (e.g. Chao, 1968; Li & Thompson, 1981). The behavior of numerals and bare nouns in (6) is similar to that in number marking languages such as English and Hindi; however, unlike English or Hindi, nouns in Lhasa Tibetan do not change form regardless of whether the numeral is singular or plural, as we can see from (6).

Although numeral constructions in Lhasa Tibetan differ from both number marking languages and classifier languages, they also exhibit the six properties of numeral constructions as observed in Jiang (2012):

³ Interestingly, if the bare noun is a human noun, Lhasa Tibetan seems to disallow it to be used in discourse-anaphoric contexts to refer to an antecedent in the preceding sentence:

(i) bu gcig dang bumo gcig khangpai nangla dug. bumo *(de) slabgrwaba red.
 boy one and girl one room’s inside-at exist girl that student be
 ‘There are one boy and one girl in the room. The girl is a student.’

(ii) bu gsum dang bumo gnyis khangpai nangla dug. bumo *(de gnyis) slabgrwaba red.
 boy three and girl two room’s inside-at exist girl that two student be
 ‘There are three boys and two girls in the room. The two girls are students.’

We will leave the data above for future study.

- (7) Six tendentially universal properties of numeral constructions
- a. Predicative
 - b. Restrictors of definites and quantifiers
 - c. Narrow scope existential reading
 - d. Long-distance scope and island-escaping ability
 - e. Generic interpretation
 - f. Lack of anaphoric use
- (Jiang, 2012)

By examining a wide range of languages, Jiang (2012, p. 74-121) observed that numeral constructions across languages exhibit six common properties regardless of whether a language is a number marking languages or a classifier language and also regardless of whether a language has overt article determiners or not. She further proposed that these six properties are potentially universal. When examining numeral constructions in Lhasa Tibetan, we also observe the same properties. First, numeral constructions in Lhasa Tibetan can be used as predicates:

- (8) Bkrashis dang Chering slabgrwaba gnyis red.
Tashi and Chering student two be
'Tashi and Tsering are two students.'

Second, numeral phrases in Lhasa Tibetan can be used as restrictors of demonstratives:

- (9) Nyimas slabgrwaba gnyis de la gzhus song.
Nyima-ERG student two that to hit PST
'Nyima hit those two students.'

Third, they can serve as arguments, receiving a narrow scope existential interpretation:

- (10) Nyimas deb gnyis nyo song.
Nyima-ERG book two buy PST
'Nyima bought two books.'

Fourth, numeral constructions in Lhasa Tibetan allow a long-distance scope interpretation and exhibit island-escaping ability:

- (11) Nyimas mi gnyis glas na gzugspo thangpo chags gi red.
Nyima-ERG person two hire if body healthy become would be
'If Nyima hires two persons, he would become healthy.' if > two; two > if

In (11), the numeral phrase *mi gnyis* 'two persons' occurs within the adjunct *if*-clause. It can receive a narrow scope interpretation within the *if*-clause, and the sentence can be paraphrased as 'for Nyima, the hiring of any two persons would be sufficient to make him become healthy'. The numeral phrase can also receive a wide scope reading out of the *if*-clause, and the sentence can be paraphrased as 'for Nyima, there are two specific persons; if these two persons are hired, he would become healthy'.

Fifth, numeral constructions in Lhasa Tibetan can receive a generic interpretation in generic sentences:

- (12) bu gsum-gyis sgrotse de bkyag thub gi red.
boy three-ERG table that lift can would be
'Three boys can lift that table.'

- (13) kushu gsum la sgor gnyis gnas gi red.
apple three to money-unit two cost would be
'Three apples cost 2 dollars.'

Sixth, numeral constructions in Lhasa Tibetan lack anaphoric use. In (14), the numeral phrase *zhimi gsum* 'three cats' cannot refer to the same phrase in the preceding sentence without the demonstrative *de*.

- (14) Nyimas zhimi gsum dang khyi gnyis nyo song.
Nyima-ERG cat three and dog two buy PST
zhimi gsum *(de) gongchenpo red.
cat three that expensive be
'Nyima bought three cats and two dogs. *(Those) three cats are expensive.'

In the next section, we will examine demonstrative phrases with and without numerals.

2.3 Demonstrative Phrases in Lhasa Tibetan

Lhasa Tibetan has three demonstratives, the proximal demonstrative *di* 'this', the distal demonstrative *de* 'that', and the far distal demonstrative *pha-gi* 'yon' (DeLancey, 2003, p. 273). In this paper, we will focus on the proximal demonstrative *di* and the distal demonstrative *de*.

Demonstratives in Lhasa Tibetan combine directly with bare nouns, appearing in the post-nominal position, and the whole phrase can freely occur in subject and object positions:

- (15) a. deb di/de gongchenpo red b. khos deb di/de bris song
book this/that expensive be he-ERG book this/that write PST
'This/That book is expensive.' 'He wrote this/that book.'

When demonstratives combine with numeral constructions in Lhasa Tibetan, they can appear in the position either preceding or following the numeral (see also DeLancey, 2003, p. 273):⁴

- (16) a. deb **de** gnyis
book that two
'those two books'
b. deb gnyis **de**
book two that
'those three books'

In the literature, scholars describe *di/de* in Lhasa Tibetan as both demonstratives and definite article determiners (e.g. DeLancey, 2001, p. 273; Denwood, 1999, p. 93). Below, we provide two diagnostic tests from Löbner's (1985) and Alexiadou, Haegeman, and Stavrou (2007) to argue that *de* and *di* in Lhasa Tibetan are demonstratives not definite article determiners.

Löbner (1985) proposed a diagnostic test to separate true definite article determiners from demonstratives:

⁴ As noted in DeLancey (2003, p. 273-275), when numerals occur in the position preceding the demonstratives, they can take the nominal suffix *-po*, which generally applies to adjectives in the modifying position:

- (i) *deb gnyi-po de*
book two-po that
'those two books' (DeLancey, 2003, p. 273)

(17) * The boy is sleeping and the boy is not sleeping.

(18) That boy is sleeping and that boy is not sleeping.

If *de* and *di* are definite determiners like *the* in English, we would expect that phrases containing them behave like these in (17) rather than these (18). However such a predication is not borne out:

(19) bu de/di deb klog song, bu de/di deb klog ma song.

boy that/this book read PST boy that/this book read NEG PST

‘That/This boy read a book; that/this boy did not read a book.’

The example in (19) shows that *de* and *di* in Lhasa Tibetan behave as demonstratives like *that* and *this* in English not definite article determiners.

Our second test to differentiate definite article determiners from demonstratives comes from Alexiadou et al. (2007). Alexiadou et al. (2007, p. 98) observed one difference between the two, namely, only the definite article determiner can be used to refer to a kind term, whereas demonstratives cannot:

(20) a. The dodo is extinct.

b. This dodo is extinct.

When applying this diagnostic test to Lhasa Tibetan, we observe that phrase with *de* and *di* simply cannot appear with kind-level predicates to denote kind, showing that *de* and *di* do not behave as a definite article determiner:

(21) *dom de/di rtsamed phyin shag

bear that/this extinct from-now PERF

Intended: ‘The bear is extinct.’

Having seen the properties of bare nouns, numeral constructions, and phrases with demonstratives in Lhasa Tibetan, in the next section, we provide an analysis of the Lhasa Tibetan facts shown in Section 2.

3. The Structure and Semantics of Lhasa Tibetan Noun Phrases

In this section, we will show that the particular properties of Lhasa Tibetan noun phrases noted in Section 2 are amenable to an analysis within the approach to nominal arguments by Chierchia (1998) and Dayal (2004), using well-established principles of interpretation.

3.1 Bare Arguments in Lhasa Tibetan

We will start with the assumptions about the denotation and the syntax of bare nominals. The system of noun phrase denotations that we adopt is given below (Chierchia, 1998; Partee, 1987):

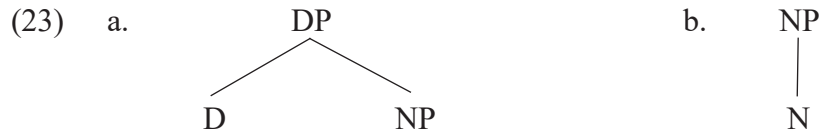
(22) a. Nominalize: $\circ P = \lambda s \iota P_s$, if $\lambda s \iota P_s$ is in K , else undefined. $\langle e, t \rangle \rightarrow \langle e^k \rangle$

b. Iota: $\iota X =$ the largest member of X if there is one, else, undefined. $\langle e, t \rangle \rightarrow \langle e \rangle$

c. Existential closure: $\exists X = \lambda P \exists y [X(y) \wedge P(y)]$ $\langle e, t \rangle \rightarrow \langle \langle e, t \rangle, t \rangle$

We adopt here the traditional view that common nouns are predicates, denoting properties. According to this view, property-denoting nouns can be turned into arguments in two ways. The first one is to argumentize nouns via the article determiner (e.g. *a/an/the* in English) (Abney, 1987; Longobardi, 1994; Szabolcsi, 1994; a.o.). As a consequence, the structure of noun phrases can be analyzed as the one in (23a), in which the functional head D(determiner) merges with an NP, forming a determiner phrase (DP). In the DP analysis of noun

phrases, Ds encode the three type-shifting operations in (22) and contribute to the different interpretations of noun phrases. Importantly, in this hypothesis, the D head can be realized in either a pronounced form or a silent form, depending on the language, and the silent/null D is only licensed in object positions and is subject to syntactic constraints (e.g. ECP) (see Longobardi, 1994).



Alternatively, the second approach turns property-denoting nouns into arguments covertly in the semantics via the type-shifters in (22) (e.g. Chierchia 1998; Dayal 2004). Consequently, the structure of noun phrases is analyzed as the one in (23b), in which bare nouns project as NPs. Importantly, in this analysis covert type-shifting operations are subject to *Blocking Principle* that favors overt article determiners over the corresponding covert type-shifting operations (24).

(24) *Blocking Principle* ('Type Shifting as Last Resort')

For any type shifting operation τ and any X: $*\tau(X)$, if there is a determiner D such that for any set X in its domain, $D(X) = \tau(X)$ (Chierchia, 1998)

The Blocking Principle is what explains the difference between the anaphoric potential of bare nominals in languages with overt definite determiners like English as opposed to languages without overt determiners like Mandarin or Hindi, for example:

- (25) a. Some children came in. *(The) children were happy. (English)
 b. *kuch baccei aaye. baccei bahut khush lage.* (Hindi)
 some children came children very happy seemed
 'Some children came. The children seemed very happy.' (Dayal, 2004)

In this paper, we adopt the second approach by Chierchia (1998) and Dayal (2004) for bare nouns in Lhasa Tibetan for two main reasons. First, as we have shown in Section 2.3, the elements *de* and *di* in Lhasa Tibetan are demonstratives rather than definite determiners, and Lhasa Tibetan does not have overt determiners.⁵ Second, as we saw in (2), bare nouns in Lhasa Tibetan can freely appear in subject and object positions with different interpretations and are not subject to any syntactic constraints, unlike bare nominals in the languages examined in Longobardi (1994). Below we will introduce two more crucial components in the approach by Chierchia (1998) and Dayal (2004) that are relevant to us.

One crucial point for the Chierchia-Dayal's approach is the difference between the indefinite readings of bare nominals and ordinary indefinites. The first allows only narrow scope indefinite readings, while the latter participates in scope interaction, as first observed by Carlson (1977) for English bare nominals:

- (26) a. Miles wants to meet policemen. want > $\exists/*\exists$ > want
 b. Miles wants to meet a policemen. want > \exists/\exists > want (Carlson, 1977)

The above difference is explained via the sort adjusting rule of Derived Kind Predication (DKP in short) by Chierchia (1998) (27) and the existential closure ' \exists ' in (22c): the former derives the obligatory narrow scope indefinite reading of kind-denoting bare nominals (26a), and the latter derives the flexible scope interpretations of ordinary indefinites (26b).

⁵ It was noted by DeLancey (2003, p. 273) that Lhasa Tibetan has an indefinite determiner *cig* which historically is a reduced form of the numeral 'one' *gcig*. According to our informants, *cig* and *gcig* are pronounced in the same way in Lhasa Tibetan and only differ in written Tibetan. Whether there are further differences between the two is unknown, and we will leave it for future study.

(27) *Derived Kind Predication (DKP):*

If P applies to objects and k denotes a kind, then $P(k) = \exists x[\cup k(x) \wedge P(x)]$

The last piece of the theory in Chierchia (1998) and Dayal (2004) relevant to our discussion is *Rank of Meaning*, which ranks the three covert type-shifters in (22), namely, kind-formation ‘ \cup ’ (22a), definite ‘ ι ’ (22b), and indefinite ‘ \exists ’ (22c) in the following way:

(28) *Ranking of Meaning:*

- a. $\cup > \{\iota, \exists\}$ (Chierchia, 1998); b. $\{\cup, \iota\} > \exists$ (revised by Dayal (2004))

In Chierchia (1998), ‘ \cup ’ ranks over ‘ ι ’ and ‘ \exists ’ (28a); this ranking is motivated by the fact that (English) plurals generally favor the kind interpretation over the indefinite one (29a). Chierchia claims that ‘ \exists ’ comes into the picture when ‘ \cup ’ is undefined (29b).

- (29) a. Machines are widespread.
b. ?? Parts of that machine are widespread. (Chierchia, 1998)

A further explanation is that kind formation ‘ \cup ’ only changes the type of its arguments without changing the information associated with it, but existential closure ‘ \exists ’ introduces quantificational force in addition to changing the type of its arguments. Kind formation ‘ \cup ’, therefore, is more meaning-preserving than ‘ \exists ’ and should get picked whenever possible. Dayal (2004), however, noted that Chierchia’s ranking in (29a) would block bare nominals in determiner-less languages from having any object-level meaning, definite or indefinite. She also noted that the same reasoning that favors ‘ \cup ’ over ‘ \exists ’ should apply to ‘ ι ’ as it also merely changes the type of its arguments without adding quantificational force. The revised ranking (28b) explains the fact that bare nominals can denote kinds as well as contextually-salient entities in languages without definite determiners. That is, definite readings are never blocked by kind formation in such languages. Ranking ‘ \exists ’ below ‘ ι ’ is based on her claim that bare nouns in such languages are *not* bona fide indefinites and that their indefinite readings are derived from their kind-level meaning:

- (30) *waimian gou mei zai-jiao.* (Chinese)
outside dog not be-barking
‘Dogs are not barking outside.’ $\neg > \exists/*\exists > \neg$ (Dayal, 2004)

With this background in place, we now illustrate how the different readings of bare nouns in Lhasa Tibetan in (2) are derived:

- (31) a. $NP_{\langle e, \iota \rangle}$
|
N
b. (2a) = extinct (\cup bears)
c. (2b) = Gen x, s [\cup dogs (x)] [clever (x, s)]
d. (2ci) (via DKP) = $\exists x$ [\cup tigers (x) \wedge see (I, x)]
e. (2cii) = ιx [tigers (x) \wedge see (I, x)]

Given that bare nouns in Lhasa Tibetan are property-referring (31a), they can be turned into arguments with a *kind* interpretation via the kind-formation operation ‘ \cup ’ (31b) due to the *Ranking of Meaning* (28b). In generic sentences (2b), the kind term provides the restriction for the generic operator *Gen* (31c) (see Krifka, 1995, for a detailed discussion on *Gen*). To derive bare nouns’ existential reading in episodic statements (2ci)/(2di), we apply DKP, as in (31d). As an immediate consequence of this, we derive the obligatory narrow scope behavior we observed in (3) in Section 2. Let us now consider the definite reading of bare nouns in (2cii)/(2dii). As we noted above, this would not follow in Chierchia’s (1998) *Ranking of Meaning* (28a), but is made possible

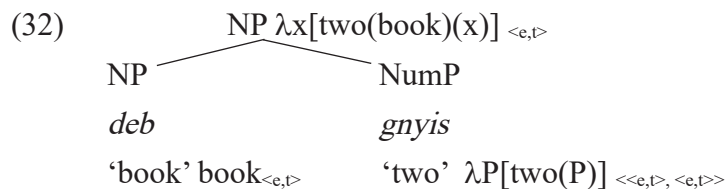
under the revision proposed by Dayal (2004) (28b). The revised *Ranking of Meaning* in (28) and lack of overt definite determiner like English *the* in Lhasa Tibetan predict that bare nouns in Lhasa Tibetan can be turned into arguments with a definite interpretation via the iota operation ‘ ι ’ (31e).

In this subsection, we saw that the kind, generic, existential, and definite interpretations of bare nouns in Lhasa Tibetan shown in Section 2 are correctly predicted by the approach of Chierchia (1998), with the specific modification of *Ranking of Meaning* from Dayal (2004). We will now turn to the interpretation and the structure of numeral constructions in Lhasa Tibetan.

3.2 The syntax and semantics of numeral constructions in Lhasa Tibetan

The behavior of Lhasa Tibetan numeral constructions is consistent with our claim that Lhasa Tibetan bare nouns denote properties. When numerals, which are property-seeking, combine with property-denoting nouns, we expect that they can combine directly, deriving the fact we presented in (6) in Section 3.1.

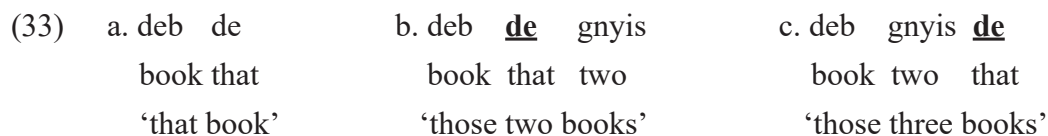
Regarding the structure of numeral construction in Lhasa Tibetan, we adopt the analysis that numerals across languages are phrasal (Borer, 2005; Di Sciullo, 2012; Ionin & Matushansky, 2006; a.o.) and place them in the adjunct position in Lhasa Tibetan (32). As for the semantic derivation of the numeral construction, we adopt the view that numerals are predicate modifiers in Ionin and Matushansky (2006). Under this view, numerals always combine with atomic predicates. In the analysis in (32), the numeral phrase *deb gnyip* ‘two book’ receives a predicative meaning, type $\langle e, t \rangle$. This analysis of numeral phrases accounts not only for the fact that they can occur in predicative positions (8) but also for the fact that they can be used as restrictors of demonstratives (9) and generic operator (12).



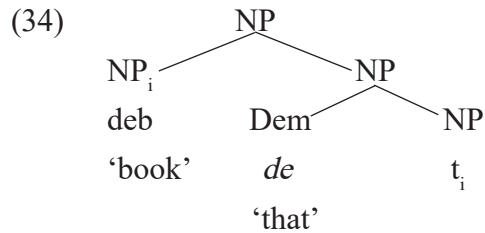
Since numeral phrases are predicative phrases in (32), we might well ask how they might be used as arguments with an indefinite interpretation, as in (10) in Section 3.1. Here we adopt the view that numerals have a predictable lexical variant in which they are indefinite determiners of type $\langle \langle e, t \rangle, \langle e, t \rangle \rangle$ (Dayal, 2012; Jiang, 2012). The resulting numeral phrase, under this analysis of numerals, is a generalized quantifier (GQ). Such an analysis immediately accounts for the fact that numeral phrases in Lhasa Tibetan are arguments (10) and show the scope behavior typical of indefinites (11).

3.3 The syntax of the demonstrative phrases in Lhasa Tibetan

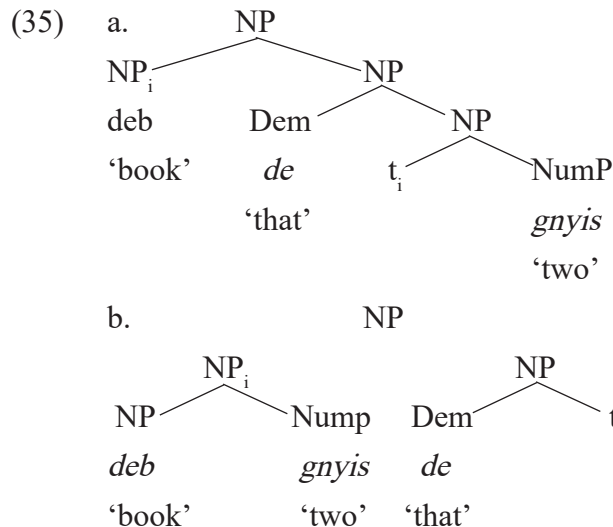
Having shown that bare nouns and numeral constructions in Lhasa Tibetan can have a predicative meaning (31)/(32), we predict that they can combine with demonstratives. Note though that there is a difference in word order in the latter case:



We adopt the view that demonstratives occur in specifier positions of NPs (Alexiadou et al., 2007; Brugè, 2000, 2002; Giusti, 1997, 2002; a.o.). We first give the structure of (33a) below:



In (34), the demonstrative appears in Spec NP position. The NP movement seen above draws on Simpson (2005, p. 309-323), where such movement is used to explain cross-linguistic variation in the ordering of constituents in the numeral classifier phrases of South East Asian languages. By adopting the NP-movement analysis, the demonstrative phrase in (33b) and (33c) have the structures in (35a) and (35b) respectively:



In (35a), the demonstrative appears in Spec NP, and the bare noun 'book' undergoes NP-fronting to the initial position of the phrase, yielding the observed word order [NP Dem Num]. In (35), the numeral modified NP *deb gnyis* 'two books' undergoes NP-fronting to the initial position, leading to the word order [NP Num Dem].

4. Conclusion and implications

This paper provided a detailed description of the syntactic distribution and semantic interpretations of bare nouns, number constructions and demonstrative phrases in Lhasa Tibetan. It was shown that bare nouns in Lhasa Tibetan share the same properties as those in classifier languages such as Mandarin. In addition to serving as predicates, Lhasa Tibetan bare nouns can serve as arguments, appearing with kind-level predicates, in generic/characterizing sentences, as well as in episodic sentences. But unlike Mandarin, numerals combine directly with bare nouns in Lhasa Tibetan in the same way as those in number marking languages like English and Hindi. However, unlike English or Hindi, nouns in Lhasa Tibetan do not change form regardless of whether the numeral is singular or plural. We further showed that numeral constructions in Lhasa Tibetan exhibit the same six properties as those in number marking languages and classifier languages as observed in Jiang (2012). We provided two diagnostic tests from Löbner's (1985) and Alexiadou et al. (2007) to argue that the so-called definite determiner *de/di* in Lhasa Tibetan are indeed demonstratives, which can appear in two different positions in numeral constructions.

We saw that the particular properties of Lhasa Tibetan noun phrases noted in this paper are amenable to an analysis within the approach to nominal arguments by Chierchia (1998) and Dayal (2004), which includes a set of well-established principles of interpretation and a principle blocking covert type shifting operations in the

presence of corresponding overt versions. Such an analysis, which is applicable to a wide range of languages, allows us to explain the properties of noun phrases in Lhasa Tibetan without making language-specific assumptions.

References

- Abney, S. (1987). *The English noun phrase in its sentential aspect*. Doctoral Dissertation, MIT.
- Alexiadou, A., Haegeman, L., & Stavrou, M. (2007). *Noun phrase in the generative perspective*. Berlin: Mouton de Gruyter.
- Borer, H. (2005). *Structuring sense, volume 1: In name only*. Oxford: Oxford University Press.
- Bošković, Z. (2005). On the locality of left branch extraction and the structure of NP. *Studia Linguistica*, 59, 1-45.
- Brugè, L. (2000). *Categorie Funzionali del nome nelle lingue Romanze*. Milan: Cisalpino.
- Brugè, L. (2002). The positions of demonstratives in the extended nominal projection. In G. Cinque (Ed.), *Functional Structure in DP and IP* (pp. 15–53). New York, NY: Oxford University Press.
- Carlson, G. (1977). *Reference to kinds in English*. Doctoral dissertation, University of Massachusetts, Amherst.
- Chao, Y. R. (1968). *A grammar of spoken Chinese*. Berkeley: University of California Press.
- Cheng, L., & Sybesma, R. (1999). Bare and not-so-bare nouns and the structure of NP, *Linguistic Inquiry*, 30 (4), 509-542.
- Chierchia, G. (1998). Reference to kinds across languages. *Natural Language Semantics*, 6, 339-405.
- Chomsky, N. (1995). *The minimalist program*. Cambridge, MA: MIT Press.
- Dayal, V. (2004). Number Marking and (in)definiteness in kind terms. *Linguistics and Philosophy*, 27 (4), 393-450.
- Dayal, V. (2012). What can South Asian classifier languages tell us about nominal systems? Paper presented at Formal Approaches to South Asian Languages (FASAL) 2. MIT.
- DeLancey, S. (2003). Lhasa Tibetan. In G. Thurgood & R. J. LaPolla (Eds.), *The Sino-Tibetan Languages* (pp. 270-288). London: Routledge.
- Denwood, P. (1999). *Tibetan*. John Benjamins.
- Di Sciullo, A.M. (2012). Asymmetry in complex numerals. Talk given at Harvard University.
- Giusti, G. (1997). The categorial status of determiners. In L. Haegeman (Ed.), *The New Comparative Syntax* (pp. 94-113). Cambridge: Cambridge University Press.
- Giusti, G. (2002). The functional structure of noun phrases: a bare phrase structure approach. In G. Cinque (Ed.), *Functional structure in DP and IP: the cartography of syntactic structures* (pp.54-90). Oxford: Oxford University.
- Ionin, T., & Matushansky, O. (2006). The composition of complex cardinals. *Journal of Semantics*, 23, 315-360.
- Jiang, L. (2012). *Nominal arguments and language variation*. Doctoral Dissertation, Harvard University.
- Krifka, M. (1995). Common nouns: A contrastive analysis of Chinese and English. In G. Carlson & F. Pelletier (Eds.), *The generic book* (pp. 398-411). Chicago: University of Chicago Press.
- Li, C., & S. Thompson. (1981). *Mandarin Chinese: A Functional Reference Grammar*. Berkeley, California: University of California Press.
- Löbner, S. (1985). Definites. *Journal of Semantics*, 4, 279–326.
- Longobardi, G. (1994). Reference and proper names: a theory of N-movement in syntax and logical form. *Linguistic Inquiry*, 25, 609-665.
- Nishi, Y. (1986). Gendai Chibetto-go hogen-no bunrui [A classification of Tibetan dialects], *Bulletin of the National Museum of Ethnology*, 11(4), 837–901.
- Partee, B. (1987). Noun phrase interpretation and type-shifting principles. In J. Groenendijk, D. de Jongh, & M. Stokhof (Eds.), *Studies in Discourse Representation Theory and the Theory of Generalized Quantifiers* (pp. 115–143). Dordrecht: Foris.
- Szabolcsi, A. (1994). The noun phrase. In F. Kiefer & K. E. Kiss (Eds), *The Syntactic Structure of Hungarian. Syntax and Semantics*, eds. (Vol. 27, pp. 179–275). New York: Academic Press.
- Shafer, R. (1955). Classification of the Sino-Tibetan languages. *Word*, 2, 94–111.

- Simpson, A. (2005). Classifiers and DP structure in Southeast Asia. In G. Cinque & R. Kayne (Eds.), *The oxford handbook of comparative syntax* (pp.806-838). Oxford: Oxford University Press.
- Wylie, T. (1959). A standard system of Tibetan transcription. *Harvard Journal of Asiatic Studies*, 22, 261-267.
- Yang, R. (2001). *Common nouns, classifiers, and quantification in Chinese*. Doctoral dissertation, Rutgers University.

A language vitality assessment for Mongolian in Inner Mongolia, China

Sarala Puthuval
University of Washington

Abstract

The paper presents a language vitality assessment for the dialects of Mongolian spoken in Inner Mongolia, China. It follows the guidelines in UNESCO's (2003) document on language vitality and endangerment, according to which a one-dimensional evaluation should not be attempted, but instead a set of nine factors should be considered separately. These include intergenerational transmission, proportion of speakers within the population, language policy, community attitudes toward the language, etc. Bringing together evidence from published sources and recent fieldwork, this assessment finds that Mongolian in Inner Mongolia scores within the endangered range on several of UNESCO's nine factors.

1. Introduction

Linguistic diversity, like biodiversity, is declining continually in the face of globalization. Despite the emergence of new urban and hybrid language varieties based on colonial languages, local and indigenous languages are gradually falling out of use as language communities become larger and more homogeneous.¹

In the state of Mongolia, Mongolian is the national language and the majority language. As such, Mongolian is not usually considered an endangered or threatened language. However, a majority of the world's Mongolian speakers live not in Mongolia but in neighboring countries such as China, Russia and Kazakhstan, where Mongolian is spoken only by small minorities within the population. For speech communities outside Mongolia, it is worth considering the possibility of Mongolian becoming threatened or endangered in those regions, especially since the dialects spoken there are significantly different from the Khalkha dialect that prevails in Mongolia.

In the People's Republic of China, Mongolian is a state-recognized minority language. The exact number of speakers is not known, but can be estimated at between two and six million, as will be discussed below. The ethnic Mongols of China are concentrated in the Inner Mongolia Autonomous Region, where 4.2 million ethnic Mongols form 17 percent of the region's population. Mongolian speakers in Inner Mongolia have a large population compared to most of the world's languages, yet it is only a small proportion of the local population. This makes Mongolian in Inner Mongolia an interesting edge case in the discussion of endangered languages. Similar cases, where there is a large population yet some reason to think the language endangered, are presented for Javanese by Ravindranath & Cohn (2014) and Northern Khmer by Vail (2006).

This paper is written with several purposes in mind. One is to contribute to the general linguistic body of knowledge about the impending loss of thousands of human languages. Another is to contribute to Mongolian linguistic studies by providing a nuanced evaluation of Mongolian's present status and future prospects within Chinese territory. A third purpose is to fill a gap in the discussion of endangered languages by examining a borderline case where language loss is at a relatively early stage, the speaker population is still large, and the

¹ Thanks to Borjigin Badma-Odsar, Yurong, and Benjamin Brosig for sharing field notes, unpublished data, and hard-to-find publications. Field research contributing to this paper was supported by the Fulbright Foundation and the University of Washington Chester Fritz and Boeing Fellowship. Research was conducted at the School for Mongolian Studies at Inner Mongolia University and at the University of Washington.

status of “endangered” is debatable. A fourth purpose is as part of an ongoing project to test the validity of different methods of endangerment evaluation by applying different methods to the same case and by repeatedly applying the same method over time (see Puthuval, 2015, 2017). The present paper is a snapshot of Mongolian’s status in Inner Mongolia circa 2015, and also a test of the evaluation framework published in 2003 by a UNESCO-convened committee of linguists (UNESCO Ad Hoc Expert Group on Endangered Languages, 2003).

UNESCO’s assessment framework is structured as a list of nine factors, or nine dimensions of language vitality, each to be evaluated separately. These nine assessments constitute the main body of this paper. Prior to that, the next sections discuss the linguistic distinctiveness of the Mongolian varieties spoken in China versus those spoken in Mongolia, and the scope of the present evaluation. The paper concludes with some comments on the strengths and weaknesses of the UNESCO assessment method, particularly the challenges of applying it to a large region rather than a local speech community.

2. Linguistic distinctiveness of Mongolian in China

Given that Mongolian is the national language of Mongolia as well as being a minority language of China, then if Mongolian disappeared from China but continued to be spoken in Mongolia, would “a language” really have been lost? There are several points of view from which one could say yes, it would.

First, there is the Ethnologue database of world languages. According to Ethnologue’s ISO 639-3 language codes, Mongolian [mon] is a macrolanguage containing two languages, Halh Mongolian [khk] spoken in Mongolia, and Peripheral Mongolian [mvf] spoken in China (<https://www.ethnologue.com/language/mon>). While some may consider this classification too rough, or too much influenced by national borders, it should still not be dismissed too easily, since the Ethnologue database is one of the very few sources available for broad quantitative studies of language endangerment (such as Ravindranath et al., 2015). In the hypothetical scenario where Mongolian disappears from China, Ethnologue’s database would register a change.

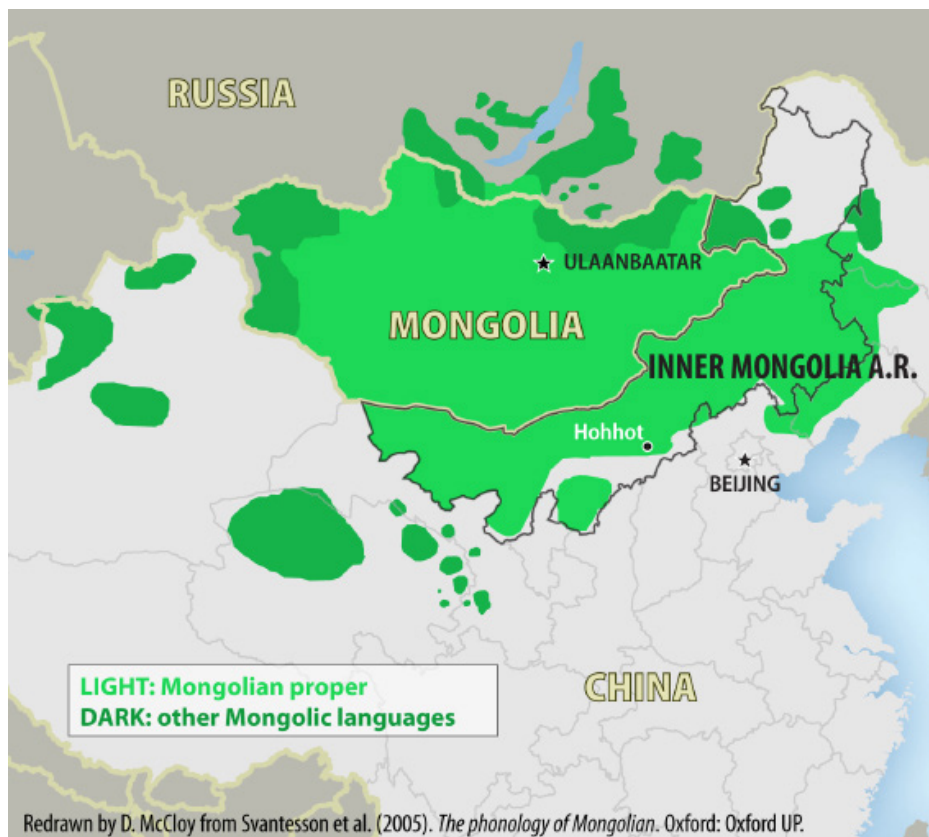


Figure 1. Geographic distribution of Mongolian and Mongolic in Northeast Asia

Next, there is the Mongolic linguistics viewpoint. Among linguists of Mongolian and Mongolic, there is a distinction drawn between *the Mongolian language*, a.k.a. *Mongolian proper*, and *the Mongolic language family*. Mongolian proper corresponds to roughly the same entity as the macrolanguage ISO 639-3 [mon], being a group of genetically related dialects descended from a hypothetical Old Mongolian. Many of the dialects (not all) are mutually intelligible; Janhunen (2012) describes Mongolian proper as a dialect continuum, which suggests that the language/dialect distinction should not be taken too literally here. The Mongolic language family, meanwhile, contains Mongolian proper plus several more distantly related languages: Mongghul, Daur, etc. The geographic distribution of Mongolian proper and the other Mongolic languages in Mongolia, China and eastern Russia is shown in Figure 1. Although there is disagreement about exactly which varieties belong to Mongolian proper and which constitute distinct Mongolic languages, the existence of Mongolian proper and the membership of certain core dialects such as Khalkha, Chakhar, and Khorchin are not disputed. Notably, the dialect diversity is much greater in Inner Mongolia than in Mongolia. In the hypothetical scenario above, if Mongolian were to disappear from China, then Mongolist linguists would report that most of the major dialects of Mongolian proper had been lost.

Finally, linguistic diversity is not the only consideration. From the point of view of cultural heritage, it would not matter even if the dialects of Mongolia and China were exactly the same linguistically. In the hypothetical scenario of Mongolian disappearing from China, the Mongolian-speaking people in China would have lost their heritage language regardless. The sustainability of Mongolian within China is relevant as long as Mongols in China care about it.

3. Scope of the present assessment

The subject of the present assessment is Mongolian in Inner Mongolia, broadly defined. This includes the official standard variety, which is based on the Chakhar dialect. It also includes any Mongolic languages, dialects or varieties spoken in the Inner Mongolia Autonomous Region of China by people who identify themselves as Mongols and/or are classified by the Chinese government as Mongols (Chinese: *měnggǔzú*; for more on the ethnic classification system in China, see the discussion of Factor 2 below.) For example, it includes speakers of the Khorchin, Chakhar, Baarin, and Kharachin-Tümed varieties, which are commonly considered dialects of Mongolian proper. It also includes speakers of the Barg, Ordos, Khamnigan, and Ejine varieties, which some authors consider dialects of Mongolian, but other authors consider to be distinct Mongolic languages. On dialect classification see e.g. Svantesson et al. (2005); Bayancogtu (2007); Janhunen (2012); IMAR Local Annals Office (2013). Most speakers of the above are classified ethnically as Mongol, with the exception of Khamnigan, some of whose speakers are classified as Evenki (Chinese: *èwēnkèzú*); they are included in this assessment to the extent that they self-identify as Mongol. The assessment does *not* concern speakers of Mongolic languages who are identified with another ethnic group. For example speakers of Daur, a Mongolic language of Northeast China, are excluded because Daur (Chinese: *dáwòěrzú*) is a distinct ethnic category for the Chinese government, and Daur speakers do not typically self-identify as Mongol.

Geographically, the assessment is limited to the territory of the Inner Mongolia Autonomous Region, plus some immediately adjacent areas in Heilongjiang, Jilin, Liaoning and Hebei Provinces. Other populations of Mongols, notably those in Xinjiang Autonomous Region and Qinghai Province, are not considered, because they are geographically not adjacent to Inner Mongolia; they are subject to different government policies; and they speak more distantly-related Mongolic varieties.²

In sum, the present assessment concerns a geographic area of about a million square miles, inhabited by over twenty million people, of which perhaps as many as four million speak any of several varieties of Mongolian or Mongolic. Thus, in contrast to many endangered language assessments, it does *not* concern

² See Limusishiden & Dede (2012) for a UNESCO-style evaluation for Mongghul, a Mongolic language spoken in Qinghai Province.

a specific local speech community. As will be seen below, this turns out to be somewhat problematic when applying UNESCO's methodology.

4. UNESCO's 2003 guidelines for assessing language vitality and endangerment

UNESCO in 2002 convened an ad hoc expert committee to address language vitality and endangerment. The committee drew up a set of guidelines for assessing the vitality or endangerment of a given language in a particular community or region. The guidelines were published as a .pdf document which is currently available on UNESCO's website (<http://www.unesco.org/new/en/culture/themes/endangered-languages/language-vitality/>). Several versions of the guidelines exist (see Bradley & Bradley, 2017), but the present assessment follows the 2003 guidelines only (UNESCO Ad Hoc Expert Group on Endangered Languages, 2003). The purpose of the guidelines is stated in the document as follows:

“Though approximately six thousand languages still exist, many are under threat. There is an imperative need for language documentation, new policy initiatives and new materials to enhance the vitality of these languages.

The cooperative efforts of language communities, language professionals, NGOs and governments will be indispensable in countering this threat. There is a pressing need to build support for language communities in their efforts to establish meaningful new roles for their endangered languages.”

Language vitality assessments, then, are presented as a policy tool for language maintenance and revitalization efforts:

“There is an urgent need in almost all countries for more reliable information about the situation of the minority languages as a basis for language support efforts at all levels.”

The present assessment, likewise, is ultimately intended to serve as a basis for language support efforts in Inner Mongolia. As long as it is published only in English, its usefulness will be limited, but translations are in the works.

The 2003 assessment guidelines are structured as a set of nine factors which are each to be evaluated on a scale from 0 to 5, where 5 is the safest and 0 is the most endangered. The nine factors are: intergenerational language transmission; the absolute number of speakers; the proportion of speakers within the total population; shifts in domains of language use; response to new domains and media; availability of materials for language education and literacy; governmental and institutional language attitudes and policies; community members' attitudes towards their own language; and the type and quality of documentation. The nine factors are further subdivided into three groups: vitality factors (1-6), attitudinal factors (7-8), and the documentation factor (9).

Each factor has its own 0-5 scale, and each is to be evaluated separately. The guidelines advise against averaging or synthesizing these nine separate evaluations into a single number. The nine factors are evaluated one by one in the following section. A summary of all nine results is provided at the end.

5. Evaluation of Mongolian according to the 9 factors

This evaluation draws on a variety of sources, including: the author's personal observations in Inner Mongolia between 2006 and 2016; a recent sociolinguistic study of intergenerational transmission of Mongolian (Puthuval, 2017); surveys of language use practices published by Mongol linguists in China; official statistics from China; and the cultural anthropology literature on language and ethnicity in Inner Mongolia.

Vitality Factors (1-6)

Vitality factors are those relating to speaker population and language use practices: intergenerational transmission, absolute population, relative population, preserving domains of language use, expanding into new domains of language use, and use of a written language.

Factor 1: Intergenerational transmission. UNESCO's criteria for intergenerational transmission are shown in Table 1. The six levels 5-0 are called *safe*, *unsafe*, *definitively endangered*, *critically endangered*, *extinct*.

Degree of endangerment	Grade	Speaker Population
safe	5	The language is used by all ages, from children up.
unsafe	4	The language is used by some children in all domains; it is used by all children in limited domains.
definitively endangered	3	The language is used mostly by the parental generation and up. The language is used mostly by the parental generation and up.
severely endangered	2	The language is used mostly by the grandparental generation and up.
critically endangered	1	The language is used by very few speakers, mostly of great-grandparental generation.
extinct	0	There are no speakers.

Reproduced from UNESCO Ad Hoc Expert Group on Endangered Languages (2003)

Table 1: Factor 1 - Intergenerational Language Transmission

From the author's observations, it is easily apparent that Mongolian is being transmitted to some children, but not all. There are very young children even today who acquire Mongolian as their first language; there are Mongolian-medium elementary schools where all subjects are taught in Mongolian; and yet there are also children (and adults of all ages) who do not speak any Mongolian even though their parents are Mongolian speakers. This would place Mongolian at either Level 4, *unsafe*, or Level 3, *definitively endangered*.

Ideally we could be more precise about *which* children are acquiring Mongolian, and *what proportion* of children. As to which children, city and town dwellers seem to experience an especially high rate of language shift, as noted by Jankowiak (2013) in the capital city of Hohhot in Western Inner Mongolia, where Mongolian is used mostly by the parental generation and up (Level 3). This was already happening in Hohhot as long ago as the 1950s (Borchigud, 1994). In Eastern Inner Mongolia, specifically the Khorchin East Middle Banner of Tongliao Prefecture, Hasierdun et al. (2012) found that Mongolian was sometimes not transmitted to children in town-dwelling families (Level 4 or 3), whereas it continued to be used by all ages in rural all-Mongol villages (Level 5). Also for Khorchin East Middle Banner, Brosig (2010-2012) corroborates these observations and adds that in the prefectural capital, Tongliao City, the transmission situation is worse than in the countryside. Still, rural communities have also experienced language shift. For example, in the Tümed banners Western Inner Mongolia, Mongolian was already extinct or nearly so (Level 0 or 1) by the 1950s (Puthuval & Wang, 2016).

As to what proportion of children, the present author's survey of some 600 ethnic Mongols from various regions, ranging in birth date from the 1920s to the 2000s, found that among ethnic Mongols whose parents spoke both Mongolian and Chinese, 84% of the children acquired Mongolian proficiently, while 16% acquired Chinese only (Puthuval, 2017). This proportion did not vary significantly for different age groups, suggesting that intergenerational loss of Mongolian is not a recent or sudden phenomenon, but rather has been going on gradually throughout the late 20th century.

Thus, particular local communities may range from Level 0 to Level 5, but most are probably at Level 3 or 4. For Inner Mongolia as a whole, Level 4 *unsafe* may be overly optimistic, since some children of Mongol

or Mongolian-speaking parents do not use Mongolian at all. The UNESCO document states that at Level 3, *definitively endangered*, some children in the community may be fluently bilingual even though others are semi-speakers or do not speak their heritage language at all. This suggests that Level 3 is an appropriate evaluation for the region.

Factor 2: Absolute number of speakers. UNESCO's guidelines provide no scale for this factor, stating instead that "It is impossible to establish a hard and fast rule for interpreting absolute numbers, but a small speech community is always at risk."

For Mongolian in Inner Mongolia, even estimating the number is a problem. The China Census reports only ethnicity data, not language data, and a significant proportion of officially registered ethnic Mongols do not speak Mongolian. Every citizen of China has an official ethnicity (Chinese: *mínzú*, also translated as "nationality") which is recorded on their ID documents, residency papers (*hùkǒuběn*) and other official documents. Ethnic identity is assigned at birth and cannot be freely altered. Official ethnic identities are drawn from a fixed list of 56 ethnic groups, including the majority Han and 55 minorities. Dual identities are not recognized; children of mixed marriages must choose the ethnicity of one parent or the other. Since ethnicity is inherited, it does not matter whether one speaks the language or identifies personally with the group. Furthermore, since members of ethnic minorities benefit from so-called "preferential policies" including easier university admission and permission to have two or more children, children of mixed Han-minority marriages nearly always select the minority ethnicity, and some people of Han descent even had themselves reclassified as Mongols when the preferential policies were first introduced (Zhao & Yang, 2009). Under the circumstances, we should expect the number of fluent Mongolian speakers to be significantly less than the official number of ethnic Mongols.

As of the 2010 census, there were 4.2 million people officially classified as ethnic Mongols in Inner Mongolia, making up 17% of the Autonomous Region's population, which is 79% Han Chinese (National Bureau of Statistics, 2010). Thus, the number of Mongolian speakers in Inner Mongolia is probably somewhere under 4.2 million. A recent volume on Mongolian dialects (IMAR Local Annals Office, 2013) estimates that there are 4.7 million Mongolian speakers in Inner Mongolia and adjacent parts of Heilongjiang, Jilin, Liaoning and Hebei (including what they call the Inner Mongolian dialects and the Barg-Buriad dialects). This estimate almost certainly errs on the high side, given what is known about the *proportion* of ethnic Mongols who speak Mongolian (see Factor 1 and Factor 3).

Even though the exact figure is not known, in terms of orders of magnitude it seems safe to say that the population of Mongolian speakers in Inner Mongolia is somewhere in the millions (10^6). It certainly does not reach the tens of millions, and almost certainly exceeds the hundreds of thousands. This estimate is sufficiently precise that we can make meaningful comparisons with the speaker populations of other languages. Mongolian in Inner Mongolia is relatively large: most of the world's languages have populations well under one million (Lewis et al., 2016). In the literature on language endangerment, a population of a few million speakers is considered large enough that population in itself is not a risk factor for the language; however, neither is it large enough to guarantee safety. Krauss (1992) tentatively proposed one million speakers as a lower limit for safety-in-numbers. Ravindranath & Cohn (2014) and Vail (2006) have argued that even languages with a million or more speakers can find themselves in an endangered state. They illustrate this with the cases of Javanese and Northern Khmer respectively, where, though the number of speakers was in the millions, the current generation of parents had lost interest in transmitting the language to children.

Factor 3: Proportion of speakers within the total population. Besides absolute population, UNESCO's guidelines also consider relative population, that is, the proportion of speakers in relation to the total population of some group to which they belong. The criteria are shown in Table 2. The criteria for the six levels are that *all*, *nearly all*, *a majority*, *a minority*, *very few*, or *none* of the reference group speak the language. The guidelines state that the reference group can be "the ethnic, religious, regional or national group with which

the speaker community identifies.” For Mongolian in Inner Mongolia, the relevant groups are ethnic and regional.

Degree of endangerment	Grade	Proportion of speakers within the total reference population
safe	5	All speak the language.
unsafe	4	Nearly all speak the language.
definitively endangered	3	A majority speak the language.
severly endangered	2	A minority speak the language.
critically endangered	1	Very few speak the language.
extinct	0	None speak the language.

Reproduced from UNESCO Ad Hoc Expert Group on Endangered Languages (2003)

Table 2: Factor 3 - Proportion of Speakers within the Total Population

For ethnic Mongols as the reference group, the literature offers several estimates, some based on educated guesses, others based on questionnaire research. In the 1980s, the Mongolian experts for the *Language Atlas of China* (Wurm et al., 1988) estimated that 14% of ethnic Mongols throughout China no longer spoke Mongolian (Dobu, 2005, 135-136). The proportion has certainly risen since then. Thirty years later, the China Language Use Survey reports that 75.52% of the ethnic Mongols surveyed throughout China were able to communicate in an “ethnic language”, and thus 24.58% could not (Xu & Dong, 2006).³ More pessimistically, Janhunen (2012) estimates that 50% of Mongols in China have lost their language. Janhunen and Dobu both remark that maintenance of Mongolian is strongest in Inner Mongolia, where Mongols are most concentrated. The present author’s field research, surveying about 600 Inner Mongolian families, found that 79% of the ethnic Mongols surveyed were able to communicate in Mongolian (Puthuval, 2017). Given the divisions on UNESCO’s scale, these various estimates all fit the criterion that “a majority” speak the language, which would place Mongolian at Level 3, *definitively endangered*.

If we take the regional population as the reference group instead of the ethnic population, the proportion of Mongolian speakers is much lower. If Inner Mongolia’s population is 17% ethnic Mongol, and ethnic Mongols are 75% Mongolian-speaking, then Inner Mongolia is only 12% Mongolian-speaking. On UNESCO’s scale this would correspond to a *minority* (Level 2, severely endangered) or perhaps *very few* (Level 1, critically endangered). Of course, a territory the size of Inner Mongolia cannot be analyzed in exactly the same way as a local speech community. Twelve percent would be a region-wide average, obscuring great variability in the concentration of Mongols in local communities. Based on 2000 census results (National Bureau of Statistics, 2004), the proportion of ethnic Mongols in individual rural townships ranges anywhere from 0%-99%. Thus there are several dozen townships with over 80% ethnic Mongols, and hundreds of townships with less than 2% Mongols in the population. This certainly has an impact on Mongolian language maintenance. For example, Ha (2008) compared all-Mongol villages against ethnically mixed villages (all in the same part of Tongliao Prefecture) and found that, in mixed villages, competence in Mongolian was declining among younger people, while in all-Mongol villages, everyone spoke Mongolian well, and furthermore the older speakers were not very competent in Chinese.

Overall, Mongolian in Inner Mongolia could be evaluated at Level 3 *definitively endangered* if the reference population is ethnic Mongols, or as low as Level 1 *critically endangered* if the reference population is

³ The China Language Survey (Xu & Dong, 2006) is primarily concerned with Standard Chinese (Putonghua) versus Chinese dialects, and therefore its reporting on minority languages is somewhat vague, using “ethnic language” (Chinese: *mínzú yǔyán*) instead of naming a specific minority language.

all residents of Inner Mongolia. Based on the author's observations, it seems reasonable to compromise at Level 2 *severely endangered*, given that most Mongolian speakers live in areas where Mongols are at least somewhat concentrated.

Factor 4: Shifts in domains of language use. Another aspect of language vitality is the domains and functions for which a language is used, for example, at home and with family members; in workplaces; in neighborhoods; as the language of instruction in schools; and so on. UNESCO's criteria for domains of language use are shown in Table 3. The scale draws a primary distinction between home domains and social (public) domains. Unlike for Factors 1 and 3, the term *endangered* is not used to label levels on this scale. Instead the level names refer specifically to the range of domains for which a language is used.

Degree of endangerment	Grade	Domains and functions
universal use	5	The language is used in all domains and for all functions.
multilingual parity	4	Two or more languages may be used in most social domains and for most functions.
dwindling domains	3	The language is used in home domains and for many functions, but the dominant language begins to penetrate even home domains.
limited or formal domains	2	The language is used in limited social domains and for several functions.
highly limited domains	1	The language is used only in a very restricted number of domains and for very few functions.
extinct	0	The language is not used in any domain for any function.

Reproduced from UNESCO Ad Hoc Expert Group on Endangered Languages (2003)

Table 3: Factor 4 - Shifts in Domains of Language Use

Factor 4 is particularly difficult to assess for Inner Mongolia as a whole, because the domains of use for Mongolian are different in each local speech community. For example, the author has witnessed communities where Mongols always speak Mongolian among themselves (apart from loanwords); communities where younger people code-switch with Chinese and older people stick to Mongolian; communities where everyone code-switches; and communities where, even in the same family, some members could speak Mongolian and others could not.

The previously cited two studies of Khorchin East Middle Banner of Tongliao Prefecture also provide some information about domains of use. One, a questionnaire-based study of several hundred households, found that Mongolian was the main language of communication in rural areas, while Chinese was used more in towns, even among Mongols (Hasierdun et al. 2012). The other, a qualitative UNESCO-style assessment, Brosig (2010-2012) found a similar pattern, where Mongolian retained most domains and functions in villages, but in larger towns, Chinese had more domains. Some town and city Mongolian speakers had a poor command of Mongolian and tended to use a mostly Chinese lexicon with Mongolian grammar. An example from Western Inner Mongolia comes from Bulag (2003), who describes families in the city of Dongsheng (capital of Ordos Prefecture) in the 1990s where the Chinese-speaking children could not communicate with their Mongolian-speaking grandparents. In those families, the parents had experienced discrimination due to their poor Chinese, and therefore avoided speaking Mongolian with their children in order that they would develop perfect Chinese.

Local communities thus range from Level 4 *multilingual parity* in places like rural Khorchin East Middle Banner, to Level 3 *dwindling domains* in small towns, to Level 2 *limited or formal domains* in places like Dongsheng. Level 1 *highly limited domains* can be seen in communities where there are no speakers left but Mongolian retains a symbolic identity function: for instance, it may be taught in school as a second

language or used in personal names, and the script may be used in decorative cultural items such as Chinggis Khaan icons (Khan, 1995).

One thing that holds true throughout Inner Mongolia is that, in certain domains, Mongolian is not an option and Chinese is a requirement. Certain necessary resources in life are controlled by non-Mongolian-speaking groups, and as a result certain necessary transactions can only be conducted in Chinese. As Bulag (2003) points out, most jobs are controlled by Chinese speakers, so Chinese is required for economic survival. Many Mongolian speakers work with non-Mongols and use exclusively Chinese in their workplace. Another example is in government administration: few government employees and Party officials speak Mongolian except at the village level, so getting necessary administrative business done usually requires Chinese at some point. A third example is higher education. Mongolian-medium schooling is available at the elementary, secondary and university levels. However, at the university level, the Mongolian track is restricted to majors relating to Mongolian history, language and culture. Studying a scientific or technical subject, or studying at a university outside Inner Mongolia, requires Chinese.

For Inner Mongolia as a whole, Level 3 *dwindling domains* seems appropriate. Chinese is used in most public domains and some homes; Mongolian is used in some homes, some informal social domains, and a few public domains.

Factor 5: Response to new domains and media. This factor refers to newer communication technologies, and also to any new area of language use that has arisen as a result of a change in society or living conditions. For some language communities, school is a new domain. For others, broadcast media and the internet are new. The previous factor, Factor 4, concerned the extent to which existing domains of the endangered language are being lost. In the words of the UNESCO document, Factor 5 concerns the extent to which speakers are “meeting the challenges of modernity with their language”, whatever those may be. For Mongolian in Inner Mongolia, important new domains are broadcast media; computers and the internet; and mobile phones. Print media and schooling are less new, and are discussed under Factors 4 and 6.

UNESCO’s criteria for the response to new domains and media are shown in Table 4. As with Factor 3, the levels are distinguished using scalar but numerically vague criteria: “all new domains”, “most new domains”, etc. As with Factor 4, the term *endangered* is not used to label levels on this scale; instead, labels refer to how *active* a presence the language has in new domains.

Degree of endangerment	Grade	Domains and functions
dynamic	5	The language is used in all new domains.
robust/active	4	The language is used in most new domains.
receptive	3	The language is used in many new domains.
coping	2	The language is used in some new domains.
minimal	1	The language is used in only a few new domains.
inactive	0	The language is not used in any new domains.

Reproduced from UNESCO Ad Hoc Expert Group on Endangered Languages (2003)

Table 4: Factor 5 - Response to New Domains and Media

Broadcast media. Mongolian channels of radio and television exist in Inner Mongolia and have existed since those technologies were introduced to the region. Some content is rebroadcast from Mongolia and some is produced locally. A portion of the broadcast time on Mongolian-language television is filled with Chinese or foreign shows dubbed into Mongolian. Mongolian speakers consume both Mongolian-language and Chinese-language media. By many reports (e.g. Lim & Ansaldo, 2016; Brosig, 2010-2012), consumption of Chinese-language media is higher.

Computers and internet. The Mongols in China write Mongolian using the traditional Uighurjin Mongol script, unlike the Mongols in Mongolia who use the Cyrillic Mongolian script. Since Cyrillic Mongolian is not legible to most Mongols in China, the Cyrillic Mongolian internet presence is not discussed here. Uighurjin Mongolian has some graphical properties that make it unique among the world’s writing systems. Like Arabic and related scripts, it uses contextual glyph variants, that is, the same letter has a different shape at the beginning, middle or end of a word. More unusually, Uighurjin Mongolian is perhaps the only modern-day script that has to be written vertically. As a result, there are significant technical barriers to using the script on devices and platforms which were originally designed for the Roman alphabet. While the script is not inherently more difficult to render on a computer than any other writing system, the fact remains that new hardware and software platforms are never designed with vertical Mongolian in mind and therefore, at each new advance, extra effort is required to make Mongolian compatible.

Mongolian users have found various solutions to this ongoing problem. For example, instead of using mass-market word-processing software like Microsoft Word, they commonly use the Uighurjin Mongolian word processor and input method editor (IME) developed by local company Menksoft (Mongolian: *möngke gal*) www.menksoft.com. Another example is that Mongolian corpus linguists have developed an ASCII romanization of the script so that they can use ordinary computational tools to process texts.

Of course, when the writing system is not available, users may simply switch languages, writing in Chinese using Chinese characters. Mongolian speakers in China who use the Internet generally interact with it primarily in Chinese, outside of a few Uighurjin Mongolian web sites such as www.holvoo.net and www.boljoo.net. In theory, romanized Uighurjin could be used online, but this strategy is quite rare compared to using Chinese characters.⁴ Brosig (2010-2012) observes that even in the heavily Mongolian-speaking Khorchin East Middle Banner, most writing is done in Chinese, and local government offices often do not own Mongolian word processing software.

Mobile phones: Text messages do not support Mongolian script. Because of the expense of long-distance calls, many people use texting instead, such that Mongolian family members communicate with each other using Chinese characters or Chinese Pinyin. Only a few well-educated speakers sometimes use romanized Uighurjin in text messaging. The situation has improved since smartphones were introduced. The popular social media platform Wechat (Chinese: *wēixìn*) allows users to send short audio messages, combining the cheapness and flexibility of texting with the ability to send the message in any language. Though Mongolian keyboard input is not supported on smartphone operating systems at present, apps do exist to convert Mongolian text to images which can then be sent via Wechat or other apps. The process is too cumbersome to use for everyday texting, but Mongolian speakers do use it for mass announcements, advertisements, holiday greetings, etc. For texting, Chinese (written in characters) is the norm. However, the situation is constantly evolving — perhaps Mongolian will be more accessible in a few years.

Returning to UNESCO’s scale, Mongolian in Inner Mongolia can be evaluated at approximately Level 3, *receptive*, where a language is used in “many” new domains. This may seem surprising since Mongolian is actually used in *all* of the domains discussed above (broadcast media, computers, and mobile phones), which might seem to imply Level 5. However, its use in each one is quite restricted compared to the dominant language, Chinese. The significance of new domains is that they tend to be vehicles for “expanding the scope and power of a dominant language at the expense of endangered languages” (UNESCO Ad Hoc Expert Group on Endangered Languages, 2003). Thus, Factor 5 is not just about the mere availability of the language in some particular domain, but about how widely used and easily used the language is in that domain. Because Mongolian’s use in new domains is relatively restricted, a lower ranking than 5 is appropriate. The distinction between Levels 2, 3, and 4 turns on “some, many, most”, which is somewhat subjective. Choosing the middle level seems a reasonable compromise.

⁴ Interestingly, Cyrillic-literate speakers from Mongolia frequently use romanizations in informal online contexts.

Factor 6: Availability of materials for language education and literacy. UNESCO’s criteria for the availability of written materials are shown in Table 5. Here, the levels are not named, but simply numbered 0-5. Mongolian in Inner Mongolia easily meets the criteria for Level 5. First, there is an established, centuries-old orthography and literacy tradition. As mentioned under Factor 5, Inner Mongolia uses the Traditional Uighurjin Mongolian script invented during the Mongol Empire, whose modern orthography was standardized in the 1970s. (For more on usage of the writing system, see Georg, 2016.) Second, there exist multiple grammars (e.g. Chinggaltai, 1963; Mongolian Language Research Center, 2005), monolingual and bilingual dictionaries (e.g. Rasidungrub, 1988; Mongolian Language Research Center, 1999), and both modern and classic literature. Literature from Mongolia, where Cyrillic writing is used instead of Uighurjin, is also published in Inner Mongolia in transliterated form. Everyday media such as school textbooks, newspapers (e.g. *Öbör monggol-un üder-un sonin*) and magazines (e.g. *Colmon*) continue to be produced. Third, Mongolian writing is used in both administration and education, even though it is not used universally in either. The large body of written materials is a point of strength for Mongolian in Inner Mongolia. Of course, how widely these materials are actually used is another matter, as touched on under Factors 4, 5 and 7.

Grade	Availability of written materials
5	There is an established orthography and a literacy tradition with grammars, dictionaries, texts, literature and everyday media. Writing in the language is used in administration and education.
4	Written materials exist and children may be exposed to the written form at school. Writing in the language is not used in administration.
3	Written materials exist and children may be exposed to the written form at school. Literacy is not promoted through print media.
2	Written materials exist, but they may only be useful for some members of the community; for others they may have a symbolic significance. Literacy education in the language is not a part of the school curriculum.
1	A practical orthography is known to the community and some material is being written.
0	No orthography is available to the community.

Reproduced from UNESCO Ad Hoc Expert Group on Endangered Languages (2003)

Table 5: Factor 6 - Availability of Materials for Language Education and Literacy

Attitudinal Factors (7-8)

The next group of factors is those relating to language attitudes and policies. Institutional attitudes and policies (Factor 7) are those held by the majority ethno-linguistic group and by the government. Community attitudes and policies (Factor 8) are those held by members of the language community, including active users of the language and also the larger “reference group” to which they belong, as discussed for Factor 3 above.

Factor 7: Institutional attitudes and policies. UNESCO’s criteria for institutional attitudes and policies are shown in Table 6. As with several of the other factors, the word *endangered* does not appear in the labels; rather, there is a specific set of labels ranging from *equal support* (Level 5) to *prohibition* (Level 0). The scale is somewhat problematic because the levels are not inherently sequential.

Degree of endangerment	Grade	Official attitudes towards language
equal support	5	All languages are protected.
differentiated support	4	Minority languages are protected primarily as the language of private domains. The use of the language is prestigious.
passive assimilation	3	No explicit policy exists for minority languages.
active assimilation	2	Government encourages assimilation to the dominant language. There is no protection for minority languages.
forced assimilation	1	The dominant language is the sole official language, while non-dominant languages are neither recognized nor protected.
prohibition	0	Minority languages are prohibited.

Reproduced from UNESCO Ad Hoc Expert Group on Endangered Languages (2003)

Table 6: Factor 7 - Governmental & Institutional Language Attitudes and Policies Including Official Status & Use

The Chinese government recognizes Mongolian as an ethnic minority language, which means speakers have the constitutional freedom to use and develop the language. At the same time, Standard Mandarin Chinese (Putonghua) is mandated as the official common language of China. Protections for Mongolian are only effective in the Inner Mongolia Autonomous Region and a few Mongol Autonomous Counties elsewhere. Use of spoken and written Standard Chinese is mandated in certain domains by law, and is the only option in many domains for practical reasons, when key players in those domains do not know and are not required to know Mongolian.

In practice, state support for the Mongolian language has included: standardizing the existing writing system; codifying a standard pronunciation; funding language documentation and description for both dialects and Standard Mongolian; funding arts, culture and broadcast media in Mongolian; and setting up Mongolian-medium schools and classrooms within the state education system. However, many Mongols consider that state support for the language has fallen short: Mongolian-medium schools are considered to provide lower quality education than Chinese-medium schools on average; graduates of Mongolian-medium schools have very limited employment options; few public officials have any command of Mongolian, making it impossible to conduct official business without knowing Chinese; and despite regulations about the use of Written Mongolian in public spaces, Mongolian signage is often illegible, error-ridden or missing. Borchigud (1995), Tümenjirgal (2008), Tsung (2014) and Han (2011) each review some successes and failures of language policy toward Mongolian.

Besides supporting Mongolian to some extent, the government in China also actively promotes assimilation to Standard Chinese via the *tuīguǎng Pǔtōnghuà* (“promulgate Standard Chinese”) suite of policies, which have been going on since the 1950s. The intent of *tuīguǎng Pǔtōnghuà* is mainly, of course, to replace the Sinitic dialects with a standard variety (Chen, 1998). Minority languages, unlike Sinitic dialects, are at least recognized as valid alternatives to Standard Chinese in some contexts; however, the explicit goal of minority education in China includes mastery of Putonghua. Minority language speakers, if they maintain their ethnic language at all, are expected to become bilingual with Putonghua (Dai & Cheng, 2007). Furthermore, the *tuīguǎng Pǔtōnghuà* policies open the way for officially sanctioned employment discrimination on the basis of language skills and accent, with Standard Chinese proficiency exams such as the *Hànyǔ Shǔipíng Kǎoshì* being required for some jobs.

Many countries in East Asia and the Pacific Rim follow the same pattern as China, according to Bradley (2007): they have overt language policies which are favorable to minority languages, but these policies may not be fully enforced, and furthermore may be accompanied by overt policies supporting linguistic assimilation.

Overall, Mongolian’s situation actually meets criteria for two disjoint levels on this scale: Level 4, *differentiated support*, and Level 2, *active assimilation*. Mongolian is protected and prestigious; at the same time, the government encourages assimilation to the dominant language, Standard Chinese. Level 3 is not a possible compromise, because its criterion is that “no explicit policy exists for minority languages”. Thus, for this factor, two levels must be selected.

Factor 8: Community attitudes and policies. UNESCO’s criteria for community attitudes and policies are shown in Table 7. As with Factor 6, there are no verbal labels for the six levels, only the numbers 0-5. As with Factors 3 and 6, the levels are distinguished using a scalar, but numerically vague, criterion: *all* community members are in favor of language maintenance, *most* community members are in favor, etc.

Grade	Community members’ attitudes towards language
5	<i>All</i> members value their language and wish to see it promoted.
4	<i>Most</i> members support language maintenance.
3	<i>Many</i> members support language maintenance; others are indifferent or may even support language loss.
2	<i>Some</i> members support language maintenance; others are indifferent or may even support language loss.
1	Only <i>a few</i> members support language maintenance; others are indifferent or may even support language loss.
0	<i>No one</i> cares if the language is lost; all prefer to use the dominant language.

Reproduced from UNESCO Ad Hoc Expert Group on Endangered Languages (2003)

Table 7: Factor 8 - Community Members’ Attitudes towards Their Own Language

While interviewing hundreds of Mongols about language maintenance in 2014 and 2015, the author encountered individuals who were extremely dedicated to language maintenance, individuals who had deliberately chosen not to pass on Mongolian to their children, individuals who regretted not acquiring Mongolian from their parents, individuals who had never learned Mongolian but were not bothered by it; in short, attitudes were all over the board Puthuval (2017). This allows us to rule out Level 5 and Level 0.

The study described above did not collect quantitative data about attitudes, but other recent surveys have done so. For example, Erdenituyaga (2013) surveyed urban-dwelling, Mongolian-proficient Mongols about their attitudes to the Mongolian language. Attitudes toward Mongolian were generally positive, including the belief that speaking Mongolian was more comfortable or easier than speaking Chinese. The China Language Use Survey’s findings likewise suggest that a majority of Inner Mongolian Mongols support language maintenance. In response to the question whether local elementary schools should have instruction in one language or two, 85% of Inner Mongolian residents (all ethnicities) supported monolingual education, and 14% of those thought the one language should be “a minority language”; meanwhile, among the 14% who supported bilingual education, 88% thought the two languages should be Putonghua plus “a minority language”.⁵ Since ethnic Mongols are about 17% of the region’s population, supposing they are proportionately represented in the survey, we can infer from these results that a majority of ethnic Mongols there are in favor of Mongolian being used in elementary education.

However, negative and indifferent attitudes have also been documented, especially in qualitative and ethnographic research. There are ethnic Mongols who genuinely believe that Mongolian language maintenance is not worth the trouble because Chinese is more useful and more modern. Bulag (2003) tells the story of a passionate language advocate who convinced his sister to send her child to a Mongolian-medium school, only to be blamed several years later for blighting the child’s future job prospects. Jankowiak (2013) observes that for

⁵ The China Language Use Survey lumps all minority languages together; see footnote 3.

many Hohhot Mongols, loss of Mongolian is simply a fact of life about which they are more or less indifferent. At the same time, there is a cultural expectation that good Mongols ought to speak Mongolian (Borchigud, 1996; Puthuval, 2017), which may cause indifferent or anti-Mongolian points of view to be underrepresented in quantitative surveys that probe only overt attitudes.

Overall, it seems fair to say that *many* or *most* support language maintenance, which would place Mongolian at Level 3 or 4. Given that there exist a significant minority who are indifferent or who support a shift to Chinese, let us provisionally select Level 3.

Documentation Factor (9)

The third group of factors comprises only one factor, Type and Quality of Documentation, which concerns the state of linguistic research and language documentation. The criteria for Factor 9 are shown in Table 8.

Nature of documentation	Grade	Language documentation
superlative	5	There are comprehensive grammars and dictionaries, extensive texts, and a constant flow of language materials. Abundant annotated high-quality audio and video recordings exist.
good	4	There is one good grammar and a number of adequate grammars, dictionaries, texts, literature and occasionally updated everyday media; adequate annotated high-quality audio and video recordings exist.
fair	3	There may be an adequate grammar or sufficient numbers of grammars, dictionaries and texts but no everyday media; audio and video recordings of varying quality or degree of annotation may exist.
fragmentary	2	There are some useful grammatical sketches, word-lists and texts useful for limited linguistic research but with inadequate coverage. Audio and video recordings of varying quality, with or without any annotation, may exist.
inadequate	1	There are only a few grammatical sketches, short word-lists and fragmentary texts. Audio and video recordings do not exist, are of unusable quality or are completely un-annotated.
undocumented	0	No material exists.

Reproduced from UNESCO Ad Hoc Expert Group on Endangered Languages (2003)

Table 8: Factor 9 - Type and Quality of Documentation

As already discussed, there exists an abundance of grammars and dictionaries for Mongolian, especially Standard Chakhar Mongolian (see references under Factor 6). In the realm of computational linguistics, Mongolian information processing research is an active field in China, building and analyzing text corpora from both modern and classical Written Mongolian (see e.g. Sechenhüü & Huashabuu, 2015). In the audio realm, the Corpus of Spoken Mongolian is a 20-hour, 80-speaker spontaneous dialogue corpus with multiple levels of annotation, including morphemes, phonemes, segments, and prosody (but no translation) (see Yurong, 2013). Audiovisual material is also continually being produced in the form of TV and radio broadcasts. While not necessarily annotated or easily available for linguistic analysis, at least some of the material is presumably being archived by the broadcasting corporations.

Standard Chakhar Mongolian and Written Mongolian may be well documented, but the other dialects of Inner Mongolia are much less so. Still, grammars, dictionaries, glossed texts, and/or instrumental phonetic studies have been published for many dialects. For a detailed review and bibliography of dialectological sources up to 2005, see Svantesson et al. (2005). There are two major holes in the documentation of non-standard dialects: the shortage of audiovisual materials, and the fragility of the publications. Most have been published locally in small print runs, on low-quality paper that degrades quickly, and are not widely archived.

A distinctive characteristic of Mongolian language documentation and linguistic research is that much of the work is conducted by Mongol researchers and published in Mongolian. This is an excellent sign for the present-day vitality of the standard language and the use of Written Mongolian. At the same time, if the language were to be lost, or even if only the writing system fell into disuse, some of the documentation would become inaccessible.

In sum, Standard Chakhar Mongolian is quite well documented, meeting UNESCO’s criteria for Level 5 *superlative* as far as written documentation is concerned, though being closer to Level 4 *good* when it comes to audiovisual materials. Other dialects spoken in Inner Mongolia are less well documented, ranging from Level 1 *inadequate* to Level 4 *good*. For Inner Mongolia as a whole, Level 4 seems an appropriate choice.

6. Summary of Mongolian’s status

Evaluations for each of the nine factors are collected side by side in Table 9 for comparison. Most fall in the middle range on the scale, suggesting that Mongolian in Inner Mongolia is neither completely safe nor in imminent danger. The UNESCO language vitality assessment is intentionally difficult to summarize: “Languages cannot be assessed simply by adding the numbers; we therefore suggest such simple addition *not be done*. Instead, the language vitality factors given above must be examined according to the purpose of the assessment.”

Factor	Rating					
1: Intergenerational Language Transmission	0	1	2	3	4	5
2: Absolute Number of Speakers	<i>A few million (all dialects combined)</i>					
3: Proportion of Speakers within the Total Population	0	1	2	3	4	5
4: Shifts in Domains of Language Use	0	1	2	3	4	5
5: Response to New Domains and Media	0	1	2	3	4	5
6: Availability of Materials for Language Education and Literacy	0	1	2	3	4	5
7: Governmental & Institutional Language Attitudes and Policies Including Official Status & Use	0	1	2	3	4	5
8: Community Members’ Attitudes towards Their Own Language	0	1	2	3	4	5
9: Type and Quality of Documentation	0	1	2	3	4	5

Table 9: Summary of evaluation for Mongolian

The main purpose of the present assessment was to investigate whether, despite its large population and relatively strong political position, Mongolian in Inner Mongolia might be endangered. The assessment has identified that Mongolian’s weakest areas are those pertaining to demographic decline, even though intergenerational transmission of the language has not been completely broken off (Factors 1 and 3). Mongolian’s status is uncertain, or a combination of weak and strong, when it comes to attitudes, policies, and domains of use (Factors 4, 5, 7 and 8). Mongolian’s strongest areas are those relating to the written tradition and language documentation (Factors 6 and 9). These relative strengths and weaknesses are not necessarily visible just from the numbers in Table 9, which is as it should be; the discussions in the preceding sections are an important part of the assessment.

A secondary purpose was to test the UNESCO assessment methodology, as seen below.

7. Concluding thoughts on the UNESCO methodology and its applicability

The present exercise has revealed both advantages and disadvantages of the UNESCO language endangerment assessment framework. One great strength of the framework is its ability to guide future research and language revitalization work. By breaking language vitality and endangerment down into multiple dimensions, the framework allows us to focus on one thing at a time, preventing a language's weakness in one dimension from being obscured or outweighed by strengths in another dimension, and vice versa. At the same time, it allows us to triangulate across different factors, noting interactions between them. This can make for more effective language revitalization strategies. Separating the dimensions can also reveal important gaps in the information available about a language's situation, and stimulate research to fill those gaps.

A surprising weakness of the UNESCO framework is that it seems to be best suited to informal, subjective evaluations on the basis of firsthand experience. The present paper has referred to external sources wherever possible, but it has often proven difficult to relate the published evidence directly to the evaluation: see, in particular, the discussions of Factor 1, Factor 3, Factor 4, Factor and Factor 8. UNESCO's scales for these factors are structured as levels along a continuum from most to least endangered, using quantitative but unspecific criteria like "most" vs. "many" vs. "some". This probably works well for smaller language communities with a few hundred to a few thousand speakers. In such cases, it would presumably be possible for a single observer or researcher to estimate this proportion quite accurately, and with sufficient precision for the six-level scale. This is not possible at the scale of Inner Mongolia. Where rigorous, comprehensive, quantitative data is not available—which it is not, for most of the data a UNESCO assessment would require—the only alternative is to sift through the published literature for anecdotes and localized studies, in order to at least represent the range of variation in the language's situation. As a consequence of its design, the UNESCO assessment framework is less accurate when applied to language communities with a large speaker population and/or a broad geographic extent. Despite some difficulties, the UNESCO assessment framework in its 2003 version (UNESCO Ad Hoc Expert Group on Endangered Languages, 2003) is a valuable and informative exercise. The author encourages others to adopt it.

References

- Bayancogtu (2007). *Nutuy-Un Ayalyun-u Sinjilel [Dialectology]*. Inner Mongolia People's Press.
- Borchigud, W. (1994). *When Is a Mongol? The Process of Learning in Inner Mongolia*. Doctoral dissertation, University of Washington, Seattle.
- Borchigud, W. (1995). The impact of urban ethnic education on modern Mongolian ethnicity, 1947-1966. In S. Harrell (Ed.) *Cultural Encounters on China's Ethnic Frontiers*, (pp. 248–277). University of Washington Press.
- Borchigud, W. (1996). Transgressing Ethnic and National Boundaries: Contemporary 'Inner Mongolian' Identities in China. In M. J. Brown (Ed.) *Negotiating Ethnicities in China and Taiwan*.
- Bradley, D. (2007). Language policy and language rights. In O. Miyooka, O. Sakiyama, & M. E. Krauss (Eds.) *The Vanishing Languages of the Pacific Rim*. Oxford ; New York: Oxford University Press.
- Bradley, D., & Bradley, M. (2017). *Language Endangerment*. Cambridge University Press.
- Brosig, B. (2010-2012). Language vitality of Khorchin (unpublished field notes).
- Bulag, U. (2003). Mongolian Ethnicity and Linguistic Anxiety in China. *American Anthropologist*, 105(4), 753–763.
- Chen, P. (1998). *Modern Chinese: History and Sociolinguistics*. Cambridge University Press.
- Chinggaltai (Ed.) (1963). *A Grammar of the Mongol Language*. New York: Frederick Ungar.
- Dai, Q., & Cheng, Y. (2007). Typology of Bilingualism and Bilingual Education in Minority Regions. In A. Feng (Ed.) *Bilingual Education in China: Practices, Policies, and Concepts*, (pp. 75–93). Multilingual Matters.
- Dobu (2005). *Dàobù Wénjí 道布文集 [Collected works of Dobu]*. 上海: 上海辞书出版社.

- Erdenituyaga (2013). *Xota-yin mongyolcud-un xelen-u obor-un tuxai sudulya [A sociolinguistic research on the language attitude of Mongolian residents in the urban area]*. Doctoral dissertation, Inner Mongolia University, Hohhot.
- Georg, S. (2016). Mongolian. In R. Sybesma, W. Behr, Y. Gu, Z. Handel, & C. J. Huang (Eds.) *Encyclopedia of Chinese Language and Linguistics*. Leiden: Brill.
- Ha, Y. (2008). Xiāngcūn měnggǔzǔ yǔyán shǐyòng xiànzhuàng yǔ biànciān — yǐ Nèi Měnggǔ T Shì cūnlúo diàochá wéi lì 乡村蒙古族语言使用现状与变迁——以内蒙古T市村落调查为例 [Language use and language shift among rural ethnic Mongols: a case study of villages in T. Shi, Inner Mongolia]. *Journal of the Second Northwest Minzu Institute (Philosophy and Social Sciences Edition)*, (04), 17–23.
- Han, E. (2011). The dog that hasn't barked: Assimilation and resistance in Inner Mongolia, China. *Asian Ethnicity*, 12(1), 55–75.
- Hasierdun, Bo, S., Ma, Z., Gao, L., Bao, S., & Tie, M. (2012). *Kēěrqín Zuǒyì Zhōngqí Měnggǔzǔ Yǔyán Shǐyòng Xiànzhuàng Jí Qí Yānbìàn 科尔沁左翼中旗蒙古族语言使用现状及其演变 [The Status Quo and Evolution of Language Use of Mongolian in Mid-Banner Left Wing of Horqin]*. 新时期中国少数民族语言使用情况研究丛书. Beijing: Commercial Press.
- IMAR Local Annals Office (Ed.) (2013). *Nèiměnggǔ zìzhìqū zhī fāngyánzhì 内蒙古自治区志方言志 [Annals of the Inner Mongolia Autonomous Region: Dialects]*, vol. Měnggǔyǔ juàn 蒙古语卷 [Mongolian]. 方志出版社.
- Janhunen, J. A. (2012). *Mongolian*. Amsterdam; Philadelphia: John Benjamins.
- Jankowiak, W. (2013). Urban Mongols. *Chinese Sociological Review*, 45(3), 53–73.
- Khan, A. (1995). Chinggis Khan: From imperial ancestor to ethnic hero. In S. Harrell (Ed.) *Cultural Encounters on China's Ethnic Frontiers*, (pp. 248–277). University of Washington Press.
- Krauss, M. (1992). The world's languages in crisis. *Language*, 68(1), 4–10.
- Lewis, M. P., Simons, G. F., & Fennig, C. D. (Eds.) (2016). *Ethnologue: Languages of the World, Nineteenth Edition*. Dallas, Texas: SIL International.
- Lim, L., & Ansaldo, U. (2016). *Languages in Contact*. Key topics in sociolinguistics. Cambridge University Press.
- Limusishiden, & Dede, K. (2012). The Mongghul experience: Consequences of language policy shortcomings. *International Journal of the Sociology of Language*, 2012(215).
- Mongolian Language Research Center (1999). *Mongyol Xitad Toli, Nemen Jasaysan Debter 蒙汉词典增订本 [Mongolian-Chinese Dictionary, Revised and Expanded Edition]*. Hohhot: Inner Mongolia University Press, 2nd ed.
- Mongolian Language Research Center (2005). *Odo Üye-Yin Monggol Xele [Modern Mongolian]*. Hohhot: Inner Mongolia People's Press, 2nd edition ed.
- National Bureau of Statistics (2004). *2000 nián Zhōngguó xiàn/qū rénkǒu pǔchá zīliáo huìbiān dàquán 2000年中国县/区人口普查资料汇编大全 [2000 China population census county-level tabulations]* (CD-ROM). All China Market Research Co., LTD.
- National Bureau of Statistics (2010). *Zhōngguó 2010 Nián Rénkǒu Pǔchá Zīliào 中国2010年人口普查资料 [2010 China Census Results]*.
- Puthuval, S. (2015). Language, identity and the moral high ground in Inner Mongolia: Practical problems for language shift research. In *Symposium on Language Shift in the Sinophone World*. Seattle.
- Puthuval, S. (2017). *Language Maintenance and Shift across Generations in Inner Mongolia*. Doctoral dissertation, University of Washington, Seattle.
- Puthuval, S., & Wang, X. (2016). Hohhot, Language of. In R. Sybesma, W. Behr, Y. Gu, Z. Handel, & C. J. Huang (Eds.) *Encyclopedia of Chinese Language and Linguistics*. Leiden: Brill.
- Rasidungrub (Ed.) (1988). *Suruyci-Yin Mongyol Xelen-u Toli [The Student's Mongolian Dictionary]*. Ulaanhad: Inner Mongolia Educational Press, 2nd edition ed.
- Ravindranath, M., & Cohn, A. C. (2014). Can a language with millions of speakers be endangered? *Journal of the Southeast Asian Linguistics Society*, 7, 64–75.
- Ravindranath, M., Cohn, A. C., & Pepinsky, T. (2015). Modeling social factors in language shift. In *New Ways of Analyzing Variation 44*. Toronto: University of Toronto.
- Sechenhüu, & Huashabuu (2015). On semantic role annotation for Mongolian corpora. *Inner Mongolia University Journal (Mongolian language edition)*, 2015(1).
- Svantesson, J.-O., Tsendina, A., Karlsson, A., & Franzén, V. (2005). *The Phonology of Mongolian*. Oxford University Press, USA.
- Tümenjirgal (2008). *Mongyol Xele Bicig-i Surcu Xereglexü Sudulxu Asayudal [Studying, Using and Researching the Mongolian Language and Orthography]*. Hohhot: Inner Mongolia Educational Press.

- Tsung, L. (2014). *Language Power and Hierarchy: Multilingual Education in China*. London; New York: Bloomsbury Academic.
- UNESCO Ad Hoc Expert Group on Endangered Languages (2003). *Language Vitality and Endangerment*. Paris: International Expert Meeting on UNESCO Programme Safeguarding of Endangered Languages.
- Vail, P. (2006). Can a language of a million speakers be endangered? Language shift and apathy among Northern Khmer speakers in Thailand. *International Journal of the Sociology of Language*, (178), 135–147.
- Wurm, S. A., Li, R., Baumann, T., Lee, M. W., Zhongguo she hui ke xue yuan, Australian Academy of the Humanities, & Australian National University. Research School of Pacific Studies. Dept. of Linguistics (1988). *Language atlas of China: Parts I and II*.
- Xu, J., & Dong, L. (Eds.) (2006). *Zhōngguó Yǔyán Wénzì Shíyòng Qíngkuàng Diào chá Zīliào 中国语言文字使用情况调查资料 [The China Language Use Survey]*. Beijing: Yuwen 语文出版社. Zhongguo yuyan wenzì diaocha ziliao.
- Yurong (2013). Corpus of Spoken Mongolian. In Rykin (Ed.) *Mongolic Languages: History and Present*, (pp. 141–143). Saint Petersburg: Nestor History.
- Zhao, L., & Yang, C. (2009). Nei Menggu shaoshu minzu renkou bianhua yu shaoshu minzu zhengce de guanxi 内蒙古少数民族人口变化与少数民族政策的关系 (The relationship between ethnicity policies and changes in the ethnic minority population of Inner Mongolia). *Journal of Inner Mongolia Normal University*, 38(5), 99–102.

The Acquisition of Azerbaijani Idioms by Second Language Learners

Gulnara Suleymanova
Michigan State University

Abstract

Learning idioms has always been very difficult for second language learners. These difficulties are also encountered in learning and teaching of Azerbaijani idioms. This article deals with the difficulties in learning Azerbaijani idioms and their learning methods by native speakers of English. The examples given in this article consist of idioms related to body parts, called *somatism*. The purpose of this article is to show some methods in teaching Azerbaijani idioms to native speakers of English.

1. Introduction

Idioms are a part of our life. They are used in all forms of discourse: in conversations, lectures, movies, radio, broadcasts, television programs, and so on. Teaching and learning idioms is one of the most difficult areas in which L2 teachers and learners are involved. It seems very difficult for L2 learners to function effectively without the knowledge of idioms. I will talk about how English native speakers can acquire learning Azerbaijani idioms more easily.

If you understand every word in a text and still fail to grasp what the text is about, chances are you are having trouble with the idioms. An *idiom* – is the assigning of a new meaning to a group of words which already have their own meaning. The interesting fact about most of these idioms is that they can easily be identified with the familiar parts of speech. *Phraseological idioms*, however, do not readily correlate with a given grammatical part of speech and require a paraphrase longer than a word (Barron, 1987; vi)

Review of various definitions of idiom:

An idiom is a group of two or more words used together in order to produce a figurative meaning. Idioms are traditionally characterized as fixed expressions whose overall meaning cannot be predicted from the sum of their constituents. The semantic denotation of the individual word is weakened and the pragmatic meaning of the multiword sequence is strengthened (Stubbs, 2007;165).

An idiom is an expression whose meaning cannot always be readily derived from the usual meaning of its constituent elements (Cooper, 1999; 233).

According to the Webster's Dictionary: "An idiom is an expression whose meaning cannot be predicted from the usual meanings of its constituent elements".

Sweet observed that "the meaning of each idiom is an isolated fact which cannot be inferred from the meaning of the words of which the idiom is made up (Sweet, 1889; 139).

Andras Balian defines the idiom as 'a phraseological unit whose meaning cannot be arrived at from the separate meanings of the constituents of the unit' (Balian, 1969; 3).

In my opinion, an idiom is a fixed expression that consists of two or more words which they have lost their original meanings and express a new idiomatic meaning. Idiomatic meaning can often be predicted by the lexical items within, because there is generally a close relationship between the literal and figurative meaning. These idioms are decomposable or analyzable, because the meanings of their parts contribute independently to their overall figurative meaning. However, there are some idioms that do not have any relationship between the literal and figurative meaning. These expressions are non-compositional, because their figurative meanings are not the functions of the meanings of their individual parts.

Some idioms can be used in a literal context or they can be used idiomatically. For example:

spill the beans.

Literal meaning: *to share beans*

Figurative meaning: *to reveal a secret*

Other idioms have no literal meaning at all, and can only be used idiomatically. For example,

until (till) kingdom come.

Means: *forever, for a very long time*

Example: *I don't want to wait **until kingdom come** for you to decide what you're doing.*

Note: 'until Kingdom come' is a phrase from a prayer in the Bible and means 'until the world ends'.

be head over heels in love

Means: *completely in love with another person*

Some idioms are completely regular and logical in their grammar and vocabulary:

big headed

Means: *be conceited; have an exaggerated sense of one's own importance or ability*

Some idioms are too difficult to guess correctly, because they have no association with the original meaning of the individual words:

make heads or tails of it

Means: *fail to understand, be quite confused about*

Two strategies could be shown in comprehension idioms by learners:

Idioms are first interpreted literally. If the meaning of an idiom does not fit the context in which the expression is situated, the native speaker searches for the idiom in a special mental idiom lexicon and then chooses the figurative meaning (Cooper, 1999).

Matlock considers that the figurative meaning of idioms is quicker to comprehend than their literal meaning. Native speakers may have access only to figurative meanings of words of particular word combinations that make up an idiom (Matlock, 1998).

Irujo (1986a) divides idioms into three groups to teach L2. In her experiment L2s are Spanish learners. In her division, idioms are characterized as: identical, similar, and different. In her opinion, if a literal meaning of the idiom is not understood, then learners will try to figure out its figurative meaning. Of course, the learner will meet with this idiom in speech, or any text. Then he or she can comprehend the meaning in the context, too.

to see eye to eye

Means: *to agree about someone or something with someone else*

to pull someone's leg

Means: *play a joke on*

Cooper shows four theories to explain how native English speakers comprehend idioms:

1) *The Idiom List Hypothesis* (Bobrow & Bell, 1973) - states that a native speaker who encounters an idiom first interprets it literally. If a literal does not fit the context in which the expression is situated, the native speaker chooses the figurative meaning.

2) *The Lexical Representation Hypothesis* (Swinney & Cuthler, 1979) – a native speaker, who encounters an idiom processes both the literal and the figurative meanings of the expression simultaneously, in which the context determines the more fitting interpretation.

3) *The Direct Access Hypothesis* (Gibbs, 1980, 1984; Schweigert, 1986) – a native speaker rarely considers the literal meaning of an idiomatic expression but instead retrieves the figurative meaning from the mental lexicon.

4) *The Composition Model* (Gibbs 1994; Tobossi & Zardon, 1995) – according to Gibbs (1984), “these data suggest that people attempt to do some decompositional analysis when understanding idiomatic phrases. When an idiom is decomposable, readers can assign independent meanings to its individual parts and will quickly recognize how these meaningful parts combine to form the overall figurative interpretation of the phrase” (Cooper, 1999; 236).

2. Difficulties learning Azerbaijani idioms

I can summarize the difficulties in learning Azerbaijani idioms as below:

1. Non – literalness

Idioms are not literal; they do not mean what they say. e.g. For example,

baş bışirmək

head + to cook

‘to cook head’

Means: *to tease or deceive someone; to tell someone something that is not true as a way of joking*

English equivalent: *to pull someone’s leg*

But we should mention that most idioms also have literal counterparts, which makes them even harder to learn. For example,

baş vurmaq

head + to hit

‘to hit head’

Means: *1) visit 2) apply and so on.*

Literal Meaning:

Divara başını vurdu.

The wall – DAT head – 3 SING POS – ACC hit – PAST 3 SING

‘He hit his head to the wall.’

Figurative Meaning:

Mən qonşuma baş vurdum.

I neighbour – 1 SING POS – DAT visit – PAST 1 SING

‘I visited my neighbour.’

2. Grammatical constraints

Göz qırpmamaq

an eye + not wink

‘not wink an eye’

Means: *not to sleep at all*

English equivalent: *not sleep a wink*

Russian equivalent: не сомкнуть глаз

Note: You cannot tell ***göz qırpdım** (slept a wink).

I winked an eye (literal) - it is not an idiom in a literal meaning.

Göz qırpmamaq means someone does not sleep the whole night. Here the idiom is used in negative. If we use the idioms in positive, the meaning changes. So, there are some Azerbaijani idioms that can only be used only in the negative. Therefore, it could be harder for learners to keep these in mind.

Note: *Göz* is used in the singular in most Azerbaijani idioms.

3. Multivalued idioms

Some Azerbaijani idioms have more than one figurative meanings. For example,

baş vurmaq

head + to hit (бить, ударить)

Means: 1. *to dive / to be plunged*; 2. *to call at someone / to drop in*; 3. *to appeal to someone / to ask something respectfully*. 4. *to apply for something*

Russian meanings: 1. *окунуться (нырять)*; 2. *навести (заходить к кому-либо)*; 3. *бить челом (почтительно обращаться и просить о чём-либо)*

4. Differences between Azerbaijani and English grammars

ayağına düşmək

foot – 3 SING POS – DAT to fall

‘to fall to somebody’s foot’

Means: *to bow to the feet of someone*

English equivalent: *fall at someone’s feet*

Russian equivalent: *падать к ногам*

Note: *At* and *feet* are used in Azerbaijani instead of *to* and *foot* in the English equivalent.

ayağı ağır adam

leg - 3 SING POS - heavy (person)

‘a heavy foot person’

Means: 1. *somebody who you meet does not bring success*, 2. *slow in movement*, 3. *a pregnant woman* English equivalent: *heavy - footed (Means: slow in movement)*

The noun comes before the adjective in Azerbaijani.

5. Historical specifics

Qazan ah etdi, **əqli başından getdi**. (Epic “Book of Dede Korkut”; XI century)

Kazan sighed and mind- 3 SING POS head – 3 SING POS – ABL went off.

‘Kazan sighed and went off his head.’

Əqli başından getmək

mind – 3 SING POS head – 2 SING POS – ABL to go off

‘go off one’s mind from head’

Means: *to be angry and act in a crazy way*

English equivalent: *to go off one’s head*

6. Cultural specifics

başında qoz sındırmaq

head – 3 SING POS – LOC a nut to break

‘to break a nut on somebody’s head’

Means: *to make life a burden to someone*

başında soğan əkmək

head – 3 SING POS – LOC an onion to plant

‘to plant an onion in somebody’s head’

Means: *to nag at someone day and night*

7. Inversions in Azerbaijani idioms

There are some inversions in Azerbaijani idioms, especially when used in poems. For example,

Qorxdum, ay aman, **yarıldı bağrım**. (literal and figure translation)

Here the idiom is **yarıldı bağrım**, is an inverted version of the idiom **bağrım yarıldı**.

bağrım yarıldı

breast + 1 SING POS break – PASSIVE VOICE

Means: *to suddenly be afraid of something*

English equivalent: *to get cold feet*

8. The role of suffix in a phrase

There are some suffixes that can change the total meaning of an idiom. For example, let’s compare these idioms:

qol qoymaq

hand + to put

‘to put hand’

Means: *to support someone in doing something*

ağzını açdırmaq

mouth + 2 SING POS + ACC to open + CAUSATIVE SUFFIX

‘to make someone’s mouth open’

Means: *to force someone to speak*

As we see in the examples, the suffix can change the meaning of the idiom. Based on this reason, learners should know the function of suffixes in Azerbaijani, too.

3. Methods and strategies for teaching Azerbaijani idioms

3.1 Teaching comprehension

Write a paragraph containing an idiom in a logical context, but omit the idiom. Ask learners to complete the sentence, paragraph with an idiom. (Irujo, 1984; 239) Example:

Nə vaxt ehtiyacın olsa, mən həmişə sənə hazırım.

Whenever you need, I always to youready – 1 SIN to be in PRES TENSE

‘I am always ready to whenever you need.’

Əl uzatmaq

hand to stretch

‘to give a hand’

Means: *to help somebody*

Completed sentence: I am always ready to give a hand whenever you need.

3.2 Comparing literal and figurative meanings

Have learners to draw sets of pictures showing both the literal and the idiomatic meaning of an idiom. It can be useful to learn and compare the literal and figurative meanings of some idioms. (Irujo, 1984; 239) For example:

Əl açmaq

hand open

‘open hand’

Literal meaning: *to open hand*

Figurative meaning: *to beg for*

Comparing idioms in Azerbaijani and English will enable learners to discover which idioms are identical, which are similar, and which are different.

1) *Identical*: which have exact English counterparts in the same meanings.

başboş

head – 3 SING POS empty

‘empty head’

Means: *foolish, stupid, silly*

English equivalent: *empty headed*

2) *Similar*: Idioms of this group have semantic counterparts in English, but are rather differently body worded. For example;

başdan – ayağa silahlanmış

head – ABL leg – DAT arm- PAST PARTICIPLE

‘armed from head to leg’

Means: *overly well-equipped or prepared*

English equivalent: *armed to the teeth*

3) *Different*: Idioms which have no counterparts in English, although their meaning can be derived from the conjoined meaning of their constituents.

ayağını yorğanına görə uzatmaq

leg-3 SING POS - ACC. coverlet-2 SING POS - DAT for to stretch

‘stretch your feet according to your coverlet’

Means: *someone should spend money due to his/her budget*

Similar English proverb: *cut your coat according to your cloth*

3.3 To make up sentences and stories with idioms.

Ask learners to make up sentences and stories using the literal meaning of the idioms and discuss the figurative meaning of the idiom.

Ana oğluna dedi: **Əlini aç**, sənə hədiyyəm var. (Used in literal meaning: *to open hand*)

Mom told her son: **Open your hand**, I have a gift for you.

Figurative meaning: *to beg for*

3.4 To make up dialogues with idioms

Ask learners to make up dialogues involving a literal misinterpretation of one or several idioms and discuss why the misinterpretation occurred (Irujo, 1986).

Turan: Sən Anarla çox görünürsən. Belə çıxır ki, onunla sevgilisən.

Sabina: **Dilini saxla**, Turan! Biz dostuq.

Turan: Dilimi necə saxlayım? Mən danışanda o hərəkət edir.

Turan: You're seen with Anar a lot. You must be in love with him.

Sabina: Hold your tongue, Turan! We are friends.

Turan: How can I hold my tongue? When I speak it is moving.

3.5 Explaining idioms in context

After teaching a text in Azerbaijani, we can identify idioms within the text and explain them to the learners. It will be easier for them to understand and remember for a long time.

3.6 Other activities

Other useful activities and strategies for teaching idioms include:

- Playing idiom charades
- Using audio-visuals to explain the meaning of idioms
- Making flash cards
- Matching activities
- Presenting idioms thematically
- Brainstorming activities
- Using idioms in conversations

4. What to pay attention to when teaching Azerbaijani idioms

Teaching idioms should:

- 1) depend on the language level of the learner
- 2) focus on repetition of the idioms
- 3) relate to the context within the instruction

Also, I want to mention that idioms similar in Azerbaijani and English are easy for learners to understand. For example:

çiyin-çiyinə

shoulder shoulder – DAT

Means: *in close proximity or cooperation*

English equivalent: *shoulder to shoulder*

Also, transparent idioms are easier to understand. For example:

kiminsə başını döndürmək (literally transparent)

somebody's head – 3 SING POS – ACC to turn

'turn somebody's head'

Means: *distracting someone by catching their attention with something attractive*

English equivalent: *turn somebody's head*

başını itirmək (metaphorically transparent)

head – 3 SING POS – ACC to lose

‘lose somebody’s head’

Means: *to become confused and agitated*English equivalent: *to lose one’s head*

Before starting to teach Azerbaijani idioms to native speakers of English, it is recommended to define the level of learners. Due to their levels, we will determine which idioms should be taught to them. Therefore, we can use three levels: **Beginning**, **Intermediate**, and **Advanced**.

1. *Beginning* – Teach the idioms which have the same meanings in English and Azerbaijani. These idioms are called *absolute equivalent idioms*. Absolute equivalent idioms have the same semantic, lexical and morphological parallelism in Azerbaijani and English are easier in comprehension idioms. This parallelism is called “interlingual factor” (Cornell, 1999; 6).
2. *Intermediate* – In this level we could teach close idioms which have the same meaning in Azerbaijani and English, but are expressed with different words.
3. *Advanced* – These learners have a greater access to vocabularies and will have little difficulties in learning Azerbaijani idioms, therefore, they can learn and understand any idioms very easily.

As a result, I can say that learning idioms is not so difficult as we think. We should know how to teach them to learners. At the beginning level; idioms can be added to the vocabulary, can be used in dialogues and stories. We can ask learners to complete sentences with idioms.

At the intermediate level we can give more difficult tasks related to idioms. They can find idioms related to the new words learned, they can find synonymous of the idioms in the text. For example,

baş (*head*)**başında olmaq**

head – 3 SING POS – LOC to be

‘to be in the head’

Means: *being in the lead*

At advanced level, learners should have access to a good idiom dictionary when needed. They will able to guess the meanings of idioms from context (Irujo, 1986).

5. Conclusion

Knowing idioms is very important when learning a language. Also, there are some difficulties in learning idioms in each language. They are also about the Azerbaijani language that native speakers of English can have when learning. In the article, the strategies were shown to teach easily Azerbaijani idioms to learners. The activities were described very largely. I hope that this article will be helpful in learning and teaching Azerbaijani idioms by native speakers of English.

References

- Adkins, P. G. (1968). Teaching idioms and figures of speech. *Modern Language Journal* 52: 148 -52
- Babkin, A. M., Baskakov, N. A., Orudzev, A. A. 1968. *Voprosi frazeologii i sostavleniia frazeologicheskix slovarei*. Baku: Izdatelstvo Akademii Nauk Azerbaidzhanskoi SSR
- Barron, (1987). Dictionary of American Idioms. vi
- Baliant, A. (1969). Sector analysis and idioms’, *Kivung*, 2(1), p.2-12
- Cooper, T. C. (1999). Processing of idioms by L2 learners of English. *TESOL, Quarterly*, 33(2): 233-262
- Cornell, A. (1999). An approach to identifying major pitfalls for learners. *IRAL*, 37(1), 1 - 11

- Feare, R. E. (1980). *Practice with Idioms*. New York and Oxford: Oxford University Press
- Irujo, S. (1984). A piece of cake: Learning and teaching idioms. *ELT Journal*, 40(3) July 1986
- Matlock, T. (1998). What is missing in research on idioms? *The American Journal of Psychology*, 11(4), 4-10
- Ragimzade, N. R. 1967. *Idiomaticheskije vyrazheniia v Azerbaidzhanskom iazike*. Baku
- Sweet, H. (1964). *The practical study of languages*, London, Oxford University Press, 1889, p. 139, reprinted in 1964, A.Makkai, 26
- Stubbs, M. (2007). Quantitative data on multi – word sequences in English. The case of the word *world*. In Hoey, M., Mahlberg, M> Stubbs & W. Teubert *Text, Discourse and Corpora*, 163 – 189. London: Continuum.
- Veliyeva, N. C. (2006). *Azerbaizhan – English – Russian Dictionary of Idioms*. Baku <http://www.dictionar.com/browse/idiom>

