E - ISSN: 2953-8203
P - ISSN: 2953-819X

YEREVAN STATE
UNIVERSITY

JOURNAL OF
IRANIAN LINGUISTICS

volume 1 | issue 1

JOURNAL OF IRANIAN LINGUISTICS

EDITOR-IN-CHIEF

Vardan Voskanian, Yerevan State University, Armenia

Volume 1 | issue 1

YEREVAN STATE UNIVERSITY
PUBLISHING HOUSE

ASSOCCIATE EDITORS

Hakob Avchyan, Yerevan State University, Armenia
Artyom Tonoyan, Yerevan State University, Armenia

EDITORIAL BOARD

Chiara Barbati, University of Pisa, Italy
Desmond Durkin-Meisterernst, Freie Universität Berlin, Germany
Jila Ghomeshi, University of Manitoba, Canada
Geoffrey Haig, University of Bamberg, Germany
Arsalan Kahnemuyipour, University of Toronto Mississauga, Canada
Simin Karimi, University of Arizona, USA
Paola Orsatti, Sapienza University of Rome, Italy
Ludwig Paul, Hamburg University, Germany
Mohammad Rasekh-Mahand, Bu-Ali Sina University, Iran
Hassan Rezai Baghbidi, Osaka University, Japan
Pollet Samvelian, Université Sorbonne Nouvelle, France
Jaffer Sheyholislami, Carleton University, Canada

E - ISSN: 2953-8203
P - ISSN: 2953-819X

JOURNAL OF IRANIAN LINGUISTICS
VOLUME 1 | ISSUE 1

CONTENT

VARDAN VOSKANIAN

Foreword **5-7**

ENRICO MORANO

Where the Demons Fell: A Manichaean Sogdian Manuscript in Sogdian Script from Mani's Book of the Giants **8-20**

HASSAN REZAI BAGHBIDI

A New Possible Etymology for the Classical Persian Particle mar **21-34**

PAOLA ORSATTI

On the Syntax of the Persian Classical Narrative Poetry: Constructions with a Past Participle in the Shāhnāme **35-65**

SALMAN ALIYARI BABOLGHANI

About šekam and šotor: The Development of the Initial Vs/šC- in Middle and New Persian **66-89**

HABIB BORJIAN

The Khonji Dialect of Lārestān **90-117**

SONGÜL GÜNDOĞDU, ARSALAN KAHNEMUYIPOUR, MARCEL DEN DIKKEN

EZAFE IN THE CONTEXT OF CPS: Evidence from Three Iranian Languages **118-141**

Mohammad Rasekh-Mahand

Mirativity in Persian **142-162**

Mohsen Mahdavi Mazdeh, Sarah Nehzati

Low Vowel Dissimulation in Mazandarani **163-183**

Foreword

I am honoured to present the inaugural volume of the *Journal of Iranian Linguistics*, dedicated to a field with a rich academic tradition that offers ever-expanding possibilities for the future. This field continues to reveal the complex intricacies of the Iranian branch of the Indo-European language family and its interactions with neighbouring languages.

Iranian linguistics, as an area of historical and comparative linguistics, encompasses a vast geographical and cultural landscape. It spans from the ancient languages such as Old Persian and Avestan to diverse modern languages like Persian, Kurdish, Balochi and Pashto, as well as the various modern Iranian dialects spoken within and outside of Iran. This field holds unique interdisciplinary value, as the Iranian languages have significantly shaped literary traditions over centuries.

As a result, Iranian linguistics is crucial not only for understanding the evolution and current state of the languages, but also for grasping the broader historical narratives of Central Asia and the Iranian plateau, the extensive region of historical Mesopotamia, Asia Minor, and beyond.

While there are many well-regarded academic journals in the broader fields of Iranian and Oriental Studies, as well as numerous reputable publications in the sphere of Linguistics, where scholars of Iranian languages are able to introduce their research, the lack of any journal solely dedicated to Iranian linguistics has left a notable gap in the field. The *Journal* aims to fill this gap by providing a dedicated platform for researchers to share their findings, foster scholarly dialogue, and expand the boundaries of knowledge within this diverse and multifaceted field.

The main goal of the *Journal* is to bring together scholars who approach this field from various perspectives, whether through structural, historical, sociolinguistic, or comparative methods. The scope of the *Journal* encompasses a broad range of topics within Iranian linguistics, including but not limited to phonology, phonetics, syntax, morphology, historical linguistics, sociolinguistics, computational linguistics, language policy, and language acquisition. The *Journal* aims to cover all language periods - Old, Middle and New Iranian.

This first volume offers contributions that reflect the *Journal*'s scope and mission. From Sogdian and Middle Persian to New Persian and modern Iranian dialects, including those of endangered varieties, these articles embody the breadth and depth of Iranian linguistics. They encircle

explorations through deep philological approaches, as well as discussions on dialectal variation and language contact.

Enrico Morano examines a Manichean Sogdian manuscript in Sogdian script from Mani's *Book of the Giants*, shedding light on the historical linguistic landscape of Central Asia. He specifically focuses on two unpublished fragments in Sogdian script from the Berlin Turfan collection, both from the same page and glassed together. These fragments contain a cosmogonic text concerning the falling of the demons/archons to the four directions of the earth, as well as part of the myth of the creation of the protoplasts by the archdemons *Šaqlūn* and *Pēsūs*.

Hassan Rezai Baghbidi offers a new possible etymology for the classical Persian particle *mar*. He conducts a comprehensive review of previous studies then posits that the particle serves as a focus marker derived through a grammaticalisation process from the Bactrian word *μαρο* [*mar*].

Paola Orsatti analyses the little-studied syntactic construction of Early and Classical New Persian which involves dependent constructions (phrases and clauses) of a verb in the form of a past participle, drawing primarily on examples from Ferdowsi's Shāhnāme, including other early poetry and prose texts.

Salman Aliyari Babolghani explores the development of the initial *Vs/šC-* in Middle and New Persian through his study of the words *šekam* and *šotor*, drawing on a broad range of materials from South-Western Iranian languages, as well as data from other linguistic sources, including contact languages of Middle Persian.

Habib Borjian presents insights into the Khonji dialect of Lārestān. His research highlights this dialect's unique phonological and grammatical features, thereby contributing to a better understanding of its historical development and contemporary usage.

In the realm of structural linguistics, ***Songül Gündoğdu, Arsalan Kahnemuyipour***, and ***Marcel den Dikken*** investigate the distribution of the *ezafe* morpheme in adnominal clauses across three Iranian languages: Persian, Northern Kurdish, and Zazaki, demonstrating that the behaviour of *ezafe* in these languages challenges the case analysis of *ezafe*, suggesting instead a compatibility with the inversion analysis of *ezafe*.

Mohammad Rasekh-Mahand demonstrates that in Persian the clitic *=hā* and the particle *ke,* both serve as mirative markers alongside their other functions. He furthermore shows that the use of the perfect form of verbs in Persian can, in certain context, operate as a mirative strategy, in addition to its primary role of signaling indirect evidentiality.

Mohsen Mahdavi Mazdeh and ***Sarah Nehzati*** examine low vowel dissimilation in Mazandarani.

As the inaugural issue, this volume also reflects the collaborative effort of numerous individuals. I would like to extend my gratitude to our editorial board, whose expertise and vision have been instrumental in shaping the direction of the *Journal*.

I am also grateful to the reviewers and contributors whose dedication and high standards of scholarship have ensured the academic rigour of this issue. And last but not least, I am particularly thankful to the assocciate editors Artyom Tonoyan and Hakob Avchyan, who have organised the complicated process of preparing and publishing of this journal and without the support of whom this project would have been impossible.

We believe that the *Journal of Iranian Linguistics* will serve as a productive platform for scholarly work in the field, significantly contributing to the growth and visibility of Iranian linguistics.

Finally, we invite all of our colleagues to join us in this endeavour, not only to explore the contributions in this inaugural volume, but also to actively engage with the J*ournal*, opening the floor for dialogue and establishing a common platform to share the results of studies and investigations on the diverse range of topics encompassed by Iranian linguistics.

Vardan Voskanian

Editor-in-Chief

Journal of Iranian Linguistics

YEREVAN STATE UNIVERSITY PUBLISHING HOUSE

JIL 1 (2024): 8-20

Journal of Iranian Linguistics

Where the Demons Fell. A Manichaean Sogdian Manuscript in Sogdian Script from Mani's *Book of the Giants*

Enrico Morano

Berlin/Turin

doi.org/10.46991/jil/2024.01.01

Abstract: Two unpublished Sogdian fragments in Sogdian Script of the Berlin Turfan collection, both from the same page and glassed together, contain a cosmogonic text on the falling of the demons/archons to the four directions of the earth and part of the myth of the creation of the protoplasts by the archdemons Šaqlūn and Pēsūs. This text will be proposed here as part of the Sogdian version of Mani's *Book of the Giants*.

Keywords: Sogdian fragments, Sogdian Script, Berlin Turfan collection, Mani, Book of the Giants

Enrico Morano
E-mail: ec.morano@gmail.com

Received: 16.04.2024
Revised: 16.07.2024
Accepted: 18.07.2024

This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License

Introduction

Two unpublished Sogdian fragments in Sogdian Script of the Berlin Turfan collection[1], both from the same page and glassed together[2], contain a cosmogonic text on the falling of the demons/archons to the four directions of the earth and part of the myth of the creation of the protoplasts by the

[1] A preliminary version of this paper was first read out at the meeting "Pre-Islamic Past of Middle Asia and Eastern Iran, dedicated to the memory of Boris Il'ič Maršak (1933-2006) and Valentin Germanovič Škoda (1951-2012)", Sankt Peterburg, Hermitage, October 23rd-25th 2013. I am very grateful to the Berlin-Brandenburgische Akademie der Wissenschaften and to the Staatsbibliothek zu Berlin - Preußischer Kulturbesitz for allowing me to study and publish these fragments. I would also thank the former Akademienvorhaben "Turfanforschung" and its staff for their kind hospitality in Berlin. I wish to thank particularly Christiane Reck, who kindly hosted me several times in her office in the Academy, always helping me with any request for manuscripts and with any kind of codicological problems.

[2] See the descritpion in Reck 2006, 111-112.

archdemons Šaqlūn and Pēsūs [3]. This text will be proposed here as part of the Sogdian version of Mani's *Book of the Giants*.[4]

The two fragments, although they do not join directly, seem to contain a running text, without interruption. This long, narrow page contains 28 lines written in a fine Sogdian script. The recto describes the falling of the demons/archons to the earth, and, for the first time in Manichaean literature, as far as I know, it is said that, depending on which of the four regions of the earth they fell, they were called by different names. Unfortunately, on two of these parts the text is corrupt, and we could not know how the ones who fell in the Western and in the Southern world (the most disquieting ones) were called[5].

So14255~So14256 [T II D II 115]

So14255/R/
/H/ ❀ {red}ʾwptʾm(n)[ty ❀][6]

/1/ rty mʾγʾz-ʾnt x[....]
/2/ (x)wʾt ZY sry sry (.)[...]
/3/ (ʾ)[w](t)ʾkh yxwstk x[yrʾnt]
/4/ ○○ rty wmʾt ky ZY γry •[7]
/5/ kʾβtʾy cy-ntr mʾγʾz-ʾntw
/6/ ʾskwʾt ○○ wmʾt ky ZY
/7/ kysynch[8] mrγty ○○ wmʾtw
/8/ ky ZY ZKw wš(k)[wpnʾ(?) 5-6]
/9/ ky ZY δxštyh [6]
/10/ ky ZY tʾry z-[8]
/11/ pγwštʾk [○○ wmʾt ky]
/12/ ZY ʾxšpʾh nyz-ʾyntw [...]
[...]?[9]

So14255/R/
/H/ The fall

/1/ and they began [...]
/2/ weak and one by one [...]
/3/ they went in separate places.
/4/ And there were some who inside the mountain
/5/ crevices began to live.
/6/ There were some who (lived)
/7/ (in) the dense forests. There were some
/8/ who the dr[y land(?) ...]
/9/ some, who [...] the plains
/10/ who the dark [...]
/11/ hidden [... there were some]
/12/ who would go out at night [...]
[...]?

---

[3] Reck 2006, 111 and 112: "Anthropogonischer Prosatext über die in die vier Himmelsrichtungen gefallenen Archonten oder Aborte".

[4] See Morano 2011, 108 "If it belonged to the *Book of Giants*, it could possibly be placed, like Zs1, in a kind of cosmogonical prologue to the book". It is thus convenient to give the text the signature Zs3 in the list of the *Book of the Giants* texts listed there.

[5] But see below the commentary on So14255/V/3/.

[6] Reading proposed by Yoshida 2008, 58.

[7] After the last word a point is written in black ink as a line-filler, or perhaps connecting the two parts of the compound?

[8] Cf. Buddh. Sogd. *kysnʾk*, "dense, luxuriant", see Henning 1940, 29 n. 1, where Yidgha *ḱesina* 'forest' is quoted".

[9] The two fragments are evidently from the same page, but they are not joining directly, see Reck 2006, 111. It is not clear whether one line is missing or not. Even if there is one missing line between the two fragments, the sentence ʾxšpʾh nyz-ʾyntw [...] / MN ʾʾwmrʾz-ty δβny(h)[....] means that some should go out at night for fear of the companions.

So14256/R/

/1/ MN ʾʾwmrʾz-ty δβny(h)[….]
/2/ ○○ rty ky nyz-ʾwr[tr …….]
/3/ ZKn tʾwʾntry ZK (š)[….]
/4/ kwnʾy ZY šy MN [……]
/5/ ʾʾsʾy ZY γrʾywy p(.)[……]
/6/ ○○ rty ʾnγty ʾβcʾn(p)[δ skwy ZY]
/7/ nβtʾy MN wy-šʾn(t)[……]
/8/ wβʾ ○○ rty [8-9]
/9/ ʾyw MN δβtyk[7-8]
/10/ wmʾtʾnt kt[wyšn(?)]
/11/ ky ZY ZKwyh (p)[ʾškyrʾn]
/12/ ʾβcʾnpδy wʾptʾntw βγy-š[t]
/13/ [ʾz]-γʾyrty wβʾnt ○○ ZY [wyšn]
/14/ ky ZY xwrsnw kyrʾn
/15/ wʾptʾntw ○○ rty pry-štʾktw
/16/ ʾz-γ(yrty) ʾkrtʾntw (○)[○ ZY wyš](n)[10]

So14256/R/

/1/ for fear of the companions.
/2/ And he who was weak[er …]
/3/ of the mightier one
/4/ would make […] and
/5/ would take him from […] and […] his body(?).
/6/ And the whole world [dry and]
/7/ wet from them[…]
/8/ was. And […]
/9/ one from another [separated(?)]
/10/ they were, if/that […]
/11/ those who fell onto the N[orthern]
/12/ world were called God[s].
/13/ And [those]
/14/ (who fell) onto the Eastern regions
/15/ were called
/16/ Messengers. [And those]

So14255/V/

/H/{blue}[❀ xwyckʾw]ʾk ❀

/1/ [ky ZY x]wrtxyz-cykw
/2/ [ʾβcʾnpδ](y) wʾptʾnt ○○ rty
/3/ [pšʾ](k)tw ʾz-γyrty wβʾn(t)
/4/ (rt)y wy-šn ky ZY ZKwy
/5/ nymyδcyk ʾβcʾnpδyh
/6/ wʾptʾntw ○○ rty MN
/7/ sʾty ptz-yry-str ZY γntʾk
/8/ [ystr ʾskw](n)t ○○ rty
/9/ [pzwkt(?)] ʾz-γyrty wβʾnt
/10/ [10]ky δymyδ
/11/ [10 ʾβc]ʾnpδyh
/12/ [wʾptʾn]tw ○○ rty cywyδ […](?)

So14255/V/

/H/ [Explanation (?)]

/1/ [who onto the W]estern
/2/ [world] fell,
/3/ were called [abortions(?)].
/4/ And those who in the
/5/ Southern world
/6/ fell were
/7/ more disquieting and
/8/ more evil than all the others. And
/9/ they were called [abortions(?)].
/10/ […] to these
/11/ […] of the w]orld
/12/ [they fe]ll(?). And thereafter […](?)

So14256/V/

/1/ [wmʾt](ʾ)ntw ○○ ZY kʾmnt δβtykw
/2/ [ʾskysʾr ʾtkwš (?) r](t)y cywyδ ʾʾz-h
/3/ [šklw]n ZY pysws
/4/ [ptymwx]s ○○ ky ZY šn
/5/ [5-6] (n)wtmy[11] wmʾtw
/6/ [6-7] pʾryk δy-wty
/7/ [7-8] ZY tmz-yrystr
/8/ [10-12] rty prwh
/9/ [6] ʾʾβrxsʾkw
/10/ [5](k)r(n)w(ʾn)cyʾ
/11/ [7-8 ○](○) rty kw δywth
/12/ [sʾr] wʾnkw wʾβw kt kw

So14256/V/

/1/ [they were(?)]. And they wanted
/2/ [to look upwards(?)] again and thereafter Āz
/3/ [clothed herself as Šaklū]n and Pēsūs.
/4/ And [to] them
/5/ […] was not hellish.
/6/ […] the other demons
/7/ […] and the most hellishly sagacious
/8/ […] and in the
/9/ […] lust
/10/ […] skill
/11/ […]. And to the demons
/12/ so spoke:

[10] Ornamental final -n filled with red ink.
[11] *(n)ɯtmy*: unknown word, perhaps *nɯ-tmy* 'not-hellish'? See DMTiii.2[2], 152a.

/13/ [xwr]sʾn sʾr nʾ tkwšδʾ	/13/ "Do not look to the [Ea]st,
/14/ [pʾrZY] prw ʾʾβrxsʾkw ʾyw	/14/ [but] with lust
/15/ δβtyky ʾprʾyw pcwz-δʾ[12]	/15/ copulate with each other
/16/ (x)[ypδ ryz šk](rδʾ) Z(Y) mnʾ	/16/ accomplish (your) desire and [bring to] me [...]

Commentary

So14255/R/

/1-12/ In this part of the text it is described how the fallen demons separated from one another and began to live in different places. For possible parallel texts in Manichaean literature see Appendix.

So14256/R/

/1/ Of the final letter only a long tail is visible. Reck 2006, 112 reads δβʾ(yš)[4] 'harm'. The final character looks more like a final -h though, and the reading δβny(h) 'fear' suits the context better.

/4-5/ kwnʾy and ʾʾsʾy are opt. 3rd sing. Here the simple optative seems to be used as a preterite, in spite of GMS §638 n.1, where it is said that it occurs only in Buddhist texts. Otherwise one could translate '(they) would make' and '(they) would take'.

/6/ rty ʾnγty ʾβcʾn(p)[δ skwy ZY] nβtʾy MN wy-šʾn(t)[.....] / wβʾ 'And the whole world [dry and] wet from them[...] was' Cf. *Kephalaia*, 92:

> 'And look, see! The Keeper of Splendour is set firm in the / great mind, in the camp above the pris/on of the bound ones, for he brings to nothing [a]ll the gloo[m] of de[ath]. An[d a] treachery came about, and an uprising! The sin abor[te]d, [it / tangled i]n with the soul. It became mixed with this light that it /expelled toward the image of the Ambassador. It went [... /in the] third firmament that is above the watch-t[ower / .] the Keeper of Splendour. From that place also it tangled in with the light. It was detached and came down / to that which is dry and that which is moist. It [fashio]ned the trees [up/on] the dry (land); but in the sea it immediat[ely] took form and / made a great uprising in the sea.' [Tr. Gardner 1995, 97.]

/8-10/ These lines are too fragmentary to allow a connected translation.

[12] On the right of the line /15/, on the outer margin two black/red points are visible on a misplaced little fragment stuck to the page.

So14256/R/11-16/ and So14255/V/1-9/ contain a description of the demons who fell into the four quarters of the world and how they were respectively called. If my reconstruction of the text is correct, those who fell into the nothern part of the world were called gods (or kings?), those who fell into the eastern part were called messengers (or angels), those who fell into the western part were called *abortions, and those who fell into the Southern world, who were more disturbing and evil than all the others, were called *offsprings of the abortions.

So14255/V/

/3/ The first incomplete word of the line at a first sight could be seen ending with [...](k)βnw. However, since just before ʾz-γyrty wβʾn(t) 'they were called' one should expect a plural of a noun, what appears as a -βn- can simply be a not perfectly written -t-. If so, the missing word could be restored as [pšʾ](k)tw, the name of the demons-abortions[13]. This, and below, l. /9/, if we emend *pzwkt, the demons-abortions' offsprings, could lead to the following description of the abortions desiring to see the Sun god again and being deceived by Āz disguised as Pēsūs."

So14256/V/

/1-2/ /1/ ○○ ZY kʾmnt δβtykw /2/ [ʾskysʾr ʾtkwš (?) 'and they wanted [to look upwards(?)] again', cf. M7800/II/5-8/ *ʾṭyy myδ[ry β](γyy) qšn /6/ wšyʾ ʾkṭwδʾ(r)n(d) (m)ʾγʾz(nd) /7/ ṭqwšʾṭ* ○○ *wyδpʾṭyy mrcync /8/ šmʾrʾ* 'and they remembered the beauty of the s[un-go]d. They began to look out (for him).'[14]

/5/ [5-6] (n)wtmy wmʾtw, 'was not hellish'. What was 'not hellish' was perhaps Šaqlūn's voice (in fact it is Āz who speaks through Šaqlūn) trying to convince the abortions to bring their offsprings to the two Archdemons. See the text of M7800/II/R/10-12/: 'and [the Enthymesis of Death = Āz] in Šaqlūn's voice g[ave comm]and [to the o]ther abortions', quoted below in the Appendix.

[13] On the distinction between pšʾk and pjwk see Sundermann 1994, 44 and text I in the Appendix below.
[14] See the full text below in the Appendix.

APPENDIX

Related texts

I.

The following texts, the first one, written in Sogdian in Manichaean script, followed by a passage from the *Liber scholiorum* by Theodor Bar Konai, describe part of the same cosmogonical event. The Sogdian text M7800/II/, in particular, seems to be very close to the text published here, and may attest that there were different Sogdian translations of the same Middle Persian work (in this case *The Book of the Giants*?). In this text one can see the use of two different words for the demons-abortions who fell to the earth after the demonesses bound in the sky aborted after seeing the androgynous Sun God naked (pšʾkṭ, or pšʾkṭ δywt) and for their offsprings (pjwkṭ).[15] Then Āz, disguised as Pēsūs and Šaqlūn (Namrāēl and Ašaqlūn in the Syriac text), convinced the demons-abortions (pšʾkṭ δywt) to bring their offsprings (pjwkṭ), 40.000 to each of them, to be devoured, so that, after having copulated, they could generate the two protoplasts.

— M7800/II/, Sogdian in Manichaean script

Hdl/ ○ wyδβʾγ cn ○ ○ pšʾqṭ δywtyy ○ . . .

/R/*1/ wnʾrʾm(yy) βr(y) xwrṭδʾrnd ○○ /2/ ʾrṭyy cʾnw* ***pšʾkṭ*** *wʾptnd /3/ ʾrṭyy mʾγʾzʾnd cn xʾ(x)ṭyy /4/ ʾʾp ʾṭyy cn wndyʾ βryy /5/ xwrṭ ʾṭyy myδ[ry β](γyy) qšn /6/ wšyʾ ʾkṭwδʾ(r)n(d) (m)ʾγʾz(nd) /7/ ṭqwšʾṭ ○○ wyδpʾṭyy mrcync /8/ šmʾrʾ ʾʾz prywyδδ ʾδwʾ /9/* ***pšʾq(ṭ δ)ywṭ*** *šqlwn ʾṭyy /10/ pysws p(t)[ymw](x)s ʾṭyy pr /11/ šklwnyy zβʾk [kw](p)ʾryk(ṭ) /12/* ***pšʾkṭ*** *sʾr (f)[rʾm]ʾy kṭ /13/ šmʾx ʾskyy sʾr nʾ /14/ ṭkwšδʾ pʾ šmʾx (s)[ʾn] /15/ xcy ○○ ywʾr kδryy š(m)[ʾ](x) /16/ šwδʾ ○ ʾṭyy nyrk ʾst(r)yc /17/ ʾpryw p(cw)zδʾ ʾṭy 1 δβṭyʾ /18/ ʾpryw [ʾβr]xsyy ryj škrδʾ /19/ zyʾnd znδʾ ʾṭyy mrṭ mrṭ* ***/V/****1/ xypδ* ***pjwqṭ*** *mnʾ sʾr /2/ ʾʾβrδʾ ʾṭyy ʾzw 1 wʾnw ʿyδc /3/ pršṭʾymkʾn kyy ʾskʾṭr /4/ pr ʾskyy sʾr ṭkwš nyʾz ny /5/ βwṭkʾ(m)[○○ ʾ](r)ṭyy wyšʾnd /6/* ***pšʾkṭ*** *(m)[wn](w f)rmʾn pṭycxšʾnd /7/ ʾṭy mʾyδ ʾkṭwδʾrnd 80 /8/ zʾr* ***pjwkṭ*** *wnyy šklwn ʾṭyy /9/ pyswsyy pyrnm(sʾ)r ʾʾγtδʾrnd /10/ ʾrṭyy wyš[ʾnd](p)ṭycxšʾnd /11/ ʾṭy 1 (p)[r 1 w](y)ʾkyy ṭγṭʾnd ○○ /12/ ʾrṭy(y)[4 β](r)ywr* ***pjwwq*** *ww /13/ šklwn xwrṭδʾrṭ ʾṭyy 4 /14/ [βr](y)wr xʾ (p)ysws ʾrṭy 1 /15/ [δβ](ṭy)ʾ ʾpr(yw) pcywznd ʾrṭy /16/ wʾnw wʾβnd qṭ mʾncyq*

[15] On the distinction see Sundermann 1994, 44 [= Sundermann 2001, 701].

/17/ šmʾrʾ kw my(š)[yy]βγw sʾr /18/ δʾryym wʾnw kʔ(ṭ)[xwn](y) ky cn /19/ (m)ʾx ʾʾjyʾṭ kww βγyštṭ

/Hdl/ Discourse on / the demons abortions (**pšʾqṭ δywtyy**):

/R/1/ They ate fruit from the forest. And when the abortions ***pšʾkṭ*** fell they began to drink water from the wells and to eat fruit /5/ from the trees, and they remembered the beauty of the s[un-go]d. They began to look out (for him). Thereupon the Enthymesis of Death, the Greed, dressed in the two abortion demons (***pšʾqṭ δywṭ***) Šaqlūn and /10/ Pēsūs, and in Šaqlūn's voice she g[ave comm]and [to the o]ther abortions ***pšʾkṭ***: "You, do not look upwards, for your e[nemy] (?) /15/ he is. But now, do go, and you, male with female ones, copulate and fulfil one with the other [lu]stful desire. Give birth to children, and one by one **/V/**1/ bring me your abortions (***pjwqṭ***), and I will make such a thing that you do not need to look upwards to the sky. /5/ And the abortions ***pšʾkṭ*** accepted the command and so they did. They brought eighty thousand abortions (***pjwkṭ***) before Šaqlūn and Pēsūs. /10/ And they received them and they entered one [by one] immediately. And Šaqlūn devoured [for]ty thousand abortions (***pjwkṭ***), and Pēsūs forty thousand. /15/ And they copulated with each other, and thus they said: "The intentional thought, we have it towards the su[n]-god, so that what will be born from us [will resemble] the gods [of sun and moon(?)]."

[Sundermann 1994, 45-46 = Sundermann 2001, 702-703].

— "He says that these daughters of Darkness were previously pregnant of their own nature, and when they beheld the attractive forms of the Messenger, their embryos aborted and fell to the earth. These ate the buds of the trees. Then the abortions took counsel together and recalled the form(s) of the Messenger that they had seen and said: 'Where is the form(s) that we saw?' And Ašaqlūn, son of the King of Darkness, said to the abortions: 'Give me your sons and daughters, and I will make for you a form like the one you saw.' They brought (them) and gave (them) to him. He ate the males, and the females he gave to <Namrāēl> his wife. Namrāēl and Ašaqlūn then united together, and she became pregnant from him and gave birth to a son, naming him Adam. She (again) became pregnant and bore a daughter, naming her Eve."

[Theodor Bar Konai, *Liber Scholiorum,* ed. A. Scher, Louvain, 1960, p. 317, transl. Reeves 1992, 192-193.

II.

Among the Manichaean Syriac texts published by Pedersen & Larsen (2013) there are some that we may perhaps consider as parts of the original *Book of the Giants* by Mani. The texts are unfortunately very damaged and fragmentary, and only a few lines are readable, often in unclear contexts. I will not discuss here in details what Pedersen & Larsen with great knowledge and philological skill have said in their extensive commentary on these texts[16], but I would like to mention here some similarities between the Berlin Syriac fragments and our Sogdian text So14255~So14256, which could perhaps corroborate the impression that the Berlin Manichaean Syriac texts are excerpts from the original *Book of the Giants* written in Syriac by Mani.

— Berlin Papyrus Collection, P 22364, Fr. 1+3 hair side right (Syriac)[17]

1 and some of them dwelled on the mountains
2 for eternal ages.
3 and because of the scent and odour
4 of the mountains they made for themselves
5 their dwelling places

In this text it is said that some of them dwelled in the mountains, and they made for themselves their dwelling places. In the Sogdian text above[18] we have a very similar expression: rty wmʾt ky ZY γry •/ kʾβtʾy cy-ntr mʾγʾz-ʾntw / ʾskwʾt 'And there were some who began to live inside the mountain crevices', and the subjects of the sentence are certainly the demons who fell to the earth.

— Berlin Papyrus Collection, P 22364, Fr. 1+3 flesh side left[19]

1 and (it was) in fear that he did
2 what he had ordered him,
3 that ruler of his

[16] Pedersen & Larsen 2013, 58 ff. and 202 ff. (Pedersen). On pp. 214 ff. there is an extensive discussion by Pedersen about Mani's *Book of the Giants* and the Berlin Papyrus Collection P22364.
[17] Pedersen & Larsen 2013, 58-59.
[18] So14255/R/4-5/.
[19] Pedersen & Larsen 2013, 71.

4 because of the fear of his companions
5 and the judjement which has been passed
6 on them by rulers

Here in the Syriac text the context is unclear. The phrase 'because of the fear of his companions' is similar, if not identical, to So14255/R/12/-So14256/R/1/: ZY ʾxšpʾh nyz-ʾyntw [...] /1/ MN ʾʾwmrʾz-ty δβny(h)[....] 'who would go out at night [...] / out of fear of the companions'. The contexts may not be the same in these two texts, but the word for 'companions' is often used both in the Qumran and in the Manichaean *Book of the Giants*[20], strengthening the possibility that this text belongs to Mani's *Book of the Giants*.

— Berlin Papyrus Collection, P 22364, Fr. 6 and 7 hair side middle[21]

3 alas, alas, woe, woe
4 because ... [
5 man those who fell into these
6 pains ... [

This text has perhaps no connection with the Sogdian text published here, but it has a striking parallel in the Sogdian text So20220/II/, the 'Lament of the Bound Rebel Stars', which may be strongly related to the Enochic literature: So20220/II/R/5/ rty wʾy wʾy ZY rxt rx /6/ šmʾxw ʾstʾrʾkt ky ZY šw /7/ prʾγt δʾrδ ZKw xypδ /8/ ʾwtʾkh ZY ʾpzʾth (r)[ty] /9/ šy kw kymy-(δ) [sʾ](r rty)[...] /10/ (mrty sʾr) [...](.)sδδh p(r)[w] /11/ r(nx M)[N c](y)wyδ pyδʾ[r] /12/ ʾw(..)[...]h ʾšmʾx [...] /13/ cwpr sʾr [...](ty)[...] /14/ zʾwrky-nw ʾxš[ʾwnh] '/5/ And woe, woe, alas, alas, you stars, you have abandoned your place and your homeland! [...] and /10/ to this and to the man you have [...] in deceit because of [...] you above [...] powerful rulership . . .'[22].

[20] Cf. Henning 1943, 68-69 [131-132], Text G, 1:'they took and imprisoned all the helpers (ʾʾwmrʾzṭ) that were in the heaven'. For the Qumran *Book of the Giants* see e.g. 4Q530 - 4QEnGiantsᵇ, Stuckenbruck 1997, 105. This passage has particular affinity with the Syriac text above. See also Pedersen & Larsen 2013, 225.
[21] Pedersen & Larsen 2013, 67.
[22] Morano 2016, 191-193.

BIBLIOGRAPHY

DMTiii.2[2] = Sims-Williams, N. & Durkin-Meisterernst, D. (2022).

Gardner, I. (1995), *The Kephalaia of the Teacher. The edited Coptic Manichaean Texts in Translation with Commentary*. (Nag Hammadi and Manichaean Studies XXXVII), Leiden-New York-Köln: E. J. Brill.

GMS = Gershevitch (1954).

Gershevitch, I. (1954), *A Grammar of Manichean Sogdian*, Oxford: Basil Blackwell [repr. 1961].

Henning, W. B. (1940), *Sogdica*, (James G. Forlong Fund, 21), London: The Royal Asiatic Society [= Henning 1977, II, pp. 1-68].

Henning, W. B. (1943), "The Book of the Giants" *Bulletin of the School of Oriental and African Studies*, vol. 11/1, pp. 52-74 [= Henning 1977, II, 115-37].

Henning, W. B. (1977), (M. Boyce & I. Gershevitch eds.) *Selected Papers*, I-II, (Acta Iranica, 2ème série, 14-15. *Hommages et Opera Minora*, V-VI), Téhéran-Liège: Bibliothèque Pahlavi.

Morano, E. (2011), "New Research on Mani's Book of Giants", in: *Der östliche Manichäismus Gattungs- und Werksgeschichte: Vorträge des Göttinger Symposiums vom 4.–5. März 2010*. Edited by Zekine Özertural and Jens Wilkens. Abhandlungen der Akademie der Wissenschaften zu Göttingen NS 17, Berlin-Boston: De Gruyter, pp. 101-111.

Morano, E. (2016), "Some New Sogdian Fragments Related to Mani's Book of Giants and the Problem of the Influence of Jewish Enochic Literature", in: M. Goff, L. T. Stuckenbruck & E. Morano (eds.), *Ancient Tales of Giants from Qumran and Turfan. Contexts, Traditions, and Influences* (Wissenschaftliche Untersuchungen Zum Neuen Testament 360), Tübingen: Mohr Siebeck, pp. 187-198.

Pedersen, N. A. & J. M. Larsen (2013), *Manichaean Texts in Syriac. First Editions, New Editions, and Studies*, (Corpus Fontium Manichaeorum. Series Syriaca I), Turnhout: Brepols.

Reck, Ch. (2006), *Mitteliranische Handschriften. Teil 1: Berliner Turfanfragmente manichäischen Inhalts in soghdischer Schrift* (VOHD 18), Stuttgart: Steiner.

Reeves, J. C. (1992), *Jewish Lore in Manichaean Cosmogony. Studies in the Book of Giants Traditions,* Cincinnati: Hebrew Union College Press.

Sims-Williams, N. & Durkin-Meisterernst, D. (2022), *Dictionary of Manichaean Sogdian and Bactrian*. Second edition, revised and enlarged by Nicholas Sims-Williams, (Corpus Fontium Manichaeorum. Dictionary of Manichaean Texts. Vol. III. Texts from Central Asia and China, part 2), Turnhout: Brepols.

Stuckenbruck, L. (1997), *The Book of Giants from Qumran.Texts, Translation, and Commentary*. (Texte und Studien zum antiken Judentum 63), Tübingen: Mohr Siebeck.

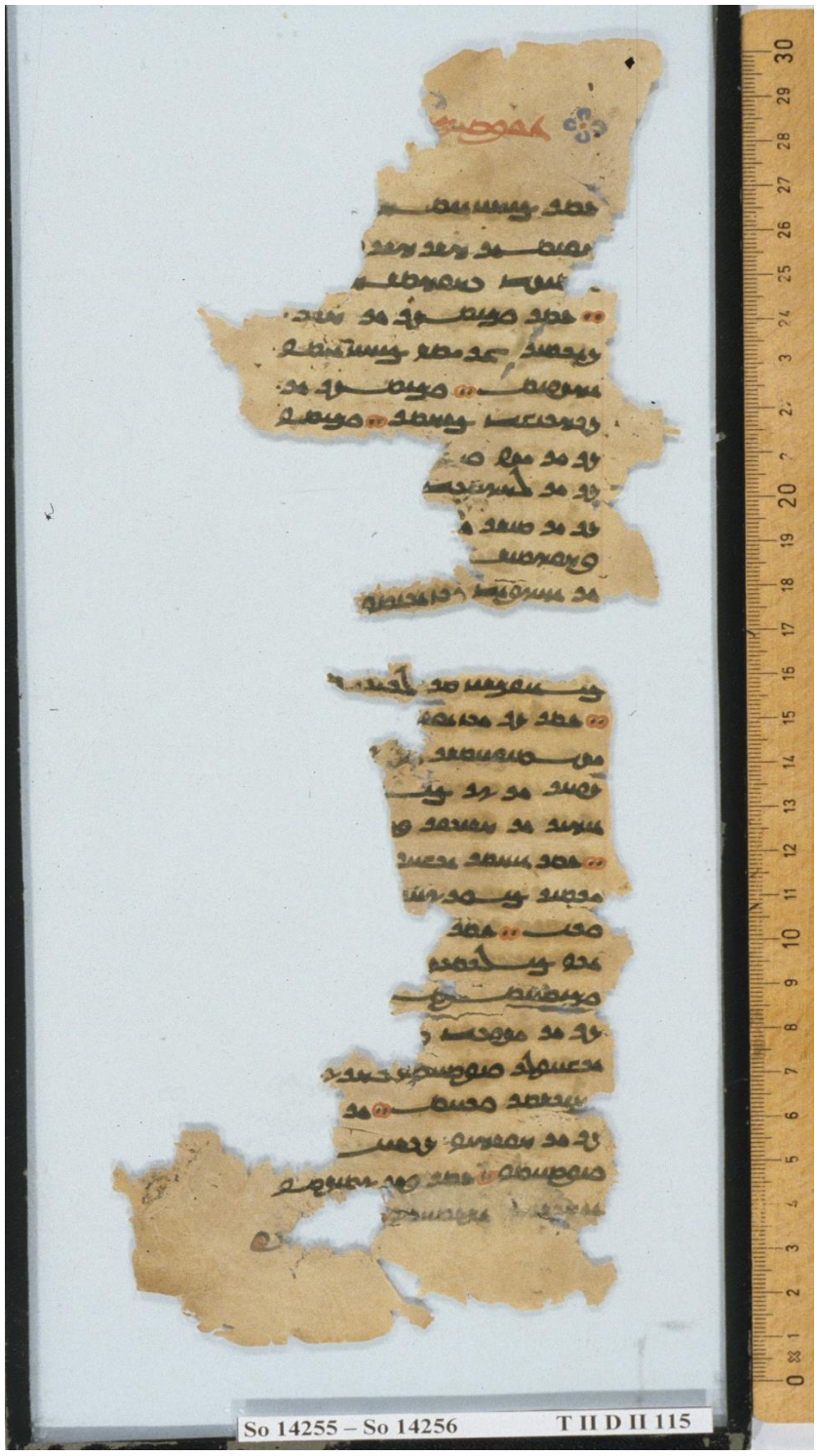
Sundermann, W. (1994), "Mani's 'Book of the Giants' and the Jewish books of Enoch: a case of terminological difference and what it implies", *Irano-Judaica* III, Jerusalem: Ben Zvi, pp. 40-48.

Sundermann, W. (2001), *Manichaica Iranica*. Ausgewählte Schriften, herausgegeben von Chr. Reck, D. Weber, C. Leurini, A. Panaino, 2 vols, (Serie Orientale Roma LXXXIX, 1) Rome: IsIAO.

Yoshida, Y. (2008), Review of Reck, Christiane, *Mitteliranische Handschriften Teil I:Berliner Turfanfragmente manichäischen Inhalts in soghdischer Schrift*, in *Indo Iranian Journal*, vol. 51, pp. 51-61.

PLATES

Fig. 1



So14255-So14256 recto

Depositum der Berlin-Brandenburgischen Akademie der Wissenschaften in der Staatsbibliothek zu Berlin – Preußischer Kulturbesitz, Orientabteilung. Photos: *Fotostelle der Staatsbibliothek zu Berlin.*

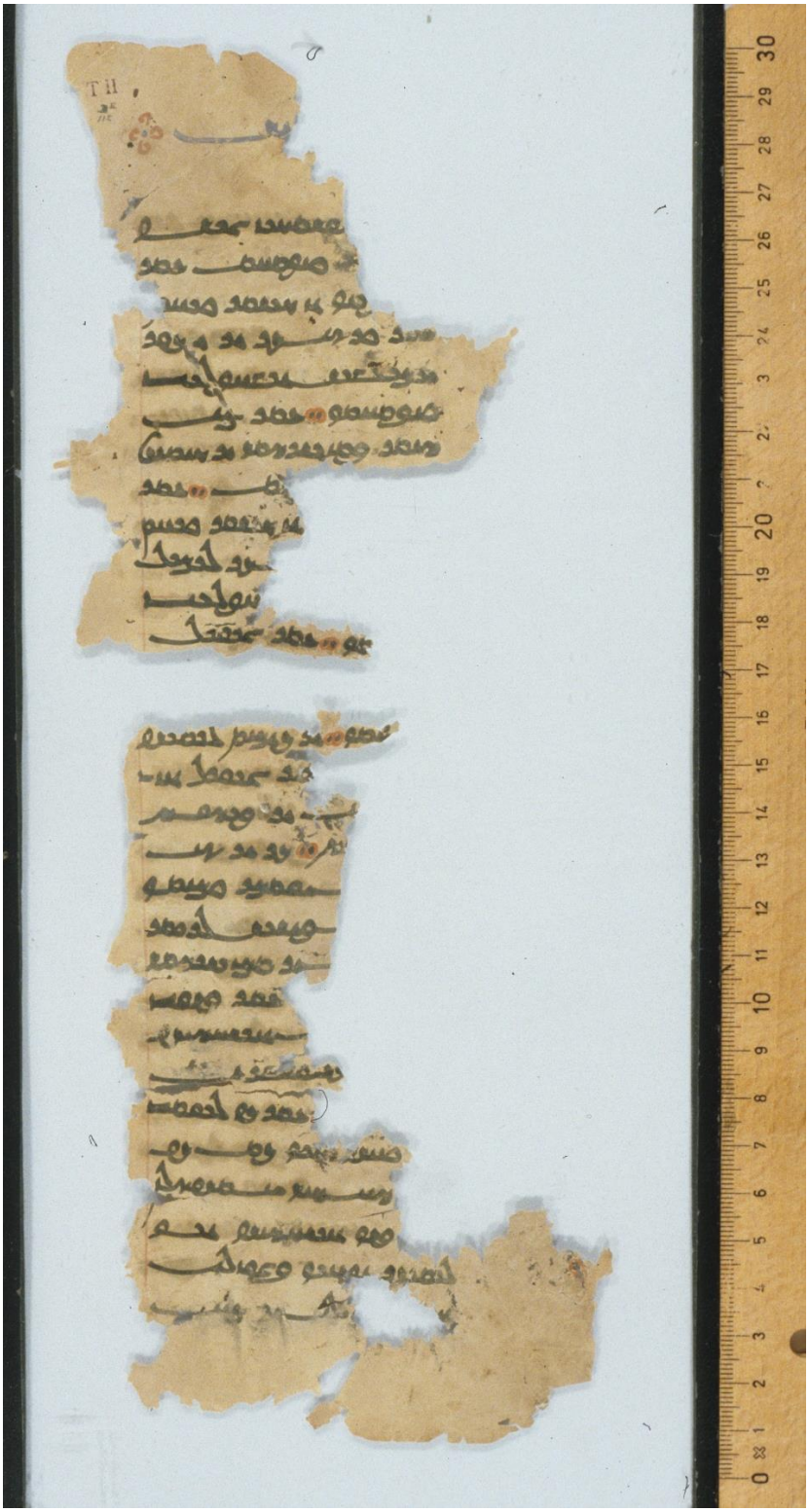


Fig. 2

So14255~So14256 verso

Depositum der Berlin-Brandenburgischen Akademie der Wissenschaften in der Staatsbibliothek zu Berlin – Preußischer Kulturbesitz, Orientabteilung. Photos: *Fotostelle der Staatsbibliothek zu Berlin.*

YEREVAN STATE UNIVERSITY PUBLISHING HOUSE

JIL 1 (2024): 21-34

Journal of Iranian Linguistics

A New Possible Etymology for the Classical Persian Particle *mar*

Hassan Rezai Baghbidi
Osaka University

doi.org/10.46991/jil/2024.01.02

Abstract: The etymology of the Classical Persian particle *mar*, which is traditionally believed to have an emphatic or restrictive sense, has long been an enigma to scholars of Iranian philology and Persian linguistics, esp. because there is no trace of it in the extant Middle Persian texts. Different etymologies have so far been proposed for *mar*, but it is often compared with the Early Judaeo-Persian preposition *azmar (i)* "for" and therefore believed to have originated from the word *mar* "number; account". The present paper reviews all previous studies and attempts to provide a new possible etymology, according to which *mar* is a focus marker derived through a grammaticalization process from the Bactrian word *μαρο [mar]* "here".

Keywords: Bactrian, Classical Persian, etymology, focus marker, grammaticalization

Hassan Rezai Baghbidi
E-mail: rezai.hmt@osaka-u.ac.jp

Received: 31.12.2023
Revised: 27.06.2024
Accepted: 30.06.2024

This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License

Introduction

According to Persian dictionaries and grammars, the Classical Persian particle *mar* serves to emphasize the subject, and is also used before the accusative (Jones 1771: 19), dative or genitive either pleonastically or in a restrictive sense (e.g., Steingass 1892: 1205; Platts 1894: 52; Phillott 1919: 57, 322; see also Darmesteter 1883: I/132, footnote 1; Horn 1898: 109-110; Gray 1937: 305; Lazard 1963: 382, 449-450). Enju Širāzi in his *Farhang-e Jahāngiri*, which he composed in India at the beginning of the 17th century, considers *mar* a pleonastic word (*az kalamāt-e zāyede*) used for the beauty of speech (*az barāye hosn-e kalām*), which sometimes conveys a restrictive meaning (*efāde-ye maʾni-ye hasr niz konad*) (ed. Afifi: I/1146). The same is repeated in the Persian dictionary *Borhān-e Qāteʿ* compiled by Moḥammad-Ḥosayn b. Ḵalaf Tabrizi in India in the middle of the 17th century (ed. Moʾin:

IV/1979). The compiler of the *Farhang-e Nezām*, the first Persian etymological dictionary, considered *mar* a pleonastic (*zāyed*) word used for decoration (*zinat*) and emphasis (*taʾkid*) (Dāʿī al-ʾIslām 1939: V/95). Similar definitions are given in other Persian dictionaries and grammars to the present day (e.g., Eṣfahāni 1872: 29; Eṣfahāni 1890: 26; Homāyunfarroḵ 1960: 1048; Maškur 1961: 190; Šafiʿi 1964: 53; Nafisi 1976: V/3235-3236; Ḵatib-e Rahbar 1988: 374; Faršidvard 2003: 451).

The origin and etymology of the particle *mar* have long been a matter of controversy. An important point which needs clarification is that there is no trace of this particle in the extant Middle Persian texts, as rightly stated by Salemann (1895: 285) and Bahār (1976: I/401). Therefore, Ṣāḥebi's (2018: 21) attempt to detect the particle *mar* in the following passage from the *Ardā Wirāz Nāmag* (Chapter 1.12-13) should be rejected. The Middle Persian word *mar* in this passage simply means "account" and has nothing to do with the particle *mar*.

> *ud pas moγ-mardān ud dastwarān ī dēn any būd hēnd,* ***az ān mar*** *andōhōmand ud purr-pīm būd hēnd*
> "Thereafter, there were other magi and religious leaders (who) were sorrowful and full of pain **on that account**" (cf. Vahman 1986: 191; Agostini 2014: 59)

As shown above, the particle *mar* has long been confused with its homonym *mar* "number; account". For example, Rückert (1854: 262) equated the particle *mar* with the Sanskrit word *mā́trā* "measure". Morgenstierne (1929: 53) proposed the probability of a connection between the particle *mar* and the Avestan root *mar* "to remember" (from the Old Iranian root **hmar* "to remember; to count", from the Indo-European root **smer* "to remember"; Cheung 2007: 137). Gray (1937: 305) tried to connect the particle *mar* with such Greek words as *μέρος* "share, portion" and *μόρος* "fate, destiny" (from the Indo-European root **smer*, Beekes 2010: II/922, 933). He then suggested that the primary meaning of the particle *mar* was "portion"; "that it was used first with the dative, and was later extended to the accusative; and that finally, coming to be felt as a mere intensive particle, it was employed even with the nominative of demonstrative pronouns as an 'empty' word" (Gray 1937: 306). Similarly, Benveniste (1938: 460) saw in the particle *mar* a specialized and quasi-prepositional function of the word *mar* "account"[1]. Gray's and Benveniste's proposals seem plausible at first sight, especially when compared with Middle Persian *az bahr ī*, Classical Persian *az bahr i* "for" (from *bahr* "portion"), and Early Judaeo-Persian *azmar (i)* "for" (from *mar* "account").

The Early Judaeo-Persian preposition *azmar*, with or without the *eẓāfe* particle *i*, is sometimes used in the sense of "because of" (Gindin 2007:

[1] une fonction spécialisée et quasi prépositionnelle du mot *mar* "compte"

III/114). It can also mark both the indirect and the direct object (see also Paul 2013: 147, 158, 163*)*:

> *ʾ**zmr y**šmwʾl*
> "**for** Šəmuʾel" (Early Karaite Document 17; Paul 2013: 148)

> *ps rwʾ hst ky gwyy ky gnd bwd ʾ**zmr** ʾn zhwmt ʾw*
> "So you may say that it was stinking **because of** its stench" (Commentary to the Book of Ezekiel 134.15; Gindin 2007: III/114)

> (*n)[ʾm](ʾ) ... ʾ**zmr**t nbyštwm*
> "I wrote (**for**) you a letter" (Private Letter 5.9; Paul 2013: 163)

> *wby hly ʾ**zmr**š pʾ dš[t]*
> "and you abandon it in a plain" (Early Argument B 13; MacKenzie 1968: 256)

Furthermore, Early Judaeo-Persian *azmar* may be complemented by the postposition *rā* to form a circumposition marking both the indirect and the direct object (cf. Shaked 2003: 210; Lazard 2009: 172):

> *ʾ**zmr** mn šmwʾl **r**ʾ*
> "**for** me, Šəmuʾel" (Early Karaite Document 4; Paul 2013: 150)

> *ʾ**zmr** ʾyn mrdwmʾn **r**ʾ nbyʾ hmy gwydšʾn*
> "**to** these people, the prophet tells them" (Commentary to the Book of Ezekiel 184.5; Gindin 2007: III/113)

> *by dʾdwm ʾ**zmr** kwn ʾw**r**ʾ*
> "I gave his blood" (Commentary to the Book of Ezekiel 132.38-133.1; Gindin 2007: III/114)

Gignoux (2010: 24) regards *az mar ī* in Middle Persian legal documents as the equivalent of Early Judaeo-Persian compound preposition *azmar (i)* "for". However, Macuch (2008: 266) has convincingly shown that *az mar ī* in Middle Persian legal documents is not a compound preposition but, rather, it simply means "from the account of, from the share of":

> ***az mar ī** man*
> "**from** my **account/share**" (Berkeley, Document 139.8; cf. Gignoux 2010: 34)

> ***az mar ī** mādar ī Farroxzād*
> "**from the account/share of** Farroxzād's mother" (Berlin, Document 19.4-5; Weber 2008: 83)

Middle Persian *az bahr ī*, Classical Persian *az bahr i*, and Early Judaeo-Persian *azmar (i)* "for" are all necessary parts of the sentence and, therefore, cannot be omitted; whereas the Classical Persian particle *mar* can always be taken out of any sentence without making it ungrammatical. That is why Lazard (1963: 451) believes that *mar* does not have any function in the structure of the sentence; rather, it seems that it highlights the word it precedes.

As a matter of fact, Horn (1893: 217) was right when he doubted the etymological connection between the particle *mar* and its homonym *mar* "number; account". Nevertheless, this hypothetical connection or the connection between the particle *mar* and the Early Judaeo-Persian preposition *azmar (i)* "for" (from *mar* "account") is still being repeated in academic books and papers (e.g., Bossong 1985: 59; Moʾayyedi and Loṭfi 2013: 111; Ṣāḥebi 2018: 22; Parizāde 2020: 237-238). Worse than that is Bahār's (1976: I/401) attempt to trace the particle *mar* back to a mark of respect and reverence, similar to the word *mār* "Lord" in Syriac Christianity!

Before discussing my own proposal as to the etymology of the particle *mar*, let us have a look at some typical examples of the use of this particle in Classical Persian texts. As it can be seen in the following examples, *mar* is used before the subject (1), the direct object with the postposition *rā* (2, 3), the direct object without the postposition *rā* (4, 5, 6), and the indirect object (i.e., the dative) with the postposition *rā* (7, 8, 9). It can also appear before a word in the genitive case with the postposition *rā* (10, 11, 12), a word in the genitive case without the postposition *rā* (13), or any other word followed by the postposition *rā* in its original meaning, i.e., "for, for the sake of, because of" (14):

(1)

پس **مر آن جفت ابداعی** صورت ابداعی بوده است

*pas **mar ān juft-i ibdāʾī** sūrat-i ibdāʾī būda ast*

"Therefore, **that innovative couple** has been an innovative form" (*Jāmiʿ al-Ḥikmatayn*, ed. Corbin and Moʿin 1953: 83)

(2)

خدای عزّ و جلّ **مر آدم را** از آن گل بیافرید

*xudāy ʾazza va jalla **mar ādam rā** az ān gil biyāfarīd*

"God – may He be honoured and glorified – created **Adam** from that clay" (*Tarjome-ye Tafsir-e Ṭabari*, ed. Yaġmāʾi 1977: II/317)

(3)

به زلیفن کردن فرستیدیم **مر ترا مر ناگرویدگان را** به دوژخ

*bi zalīfan kardan firistīdīm **mar tu rā mar nāgiravīdagān rā** ba dūžax*

"We sent **you** to warn **the unbelievers** to Hell" (*Tafsir-e Qorʾān-e Pāk*, ed. Minovi 1969: 85)

(4)

جاحظ **مر این خبر** به بعضی از کتبهای خود یاد کند

*jāhiz **mar ēn xabar** bi baʾzī az kutubhā-yi xʷad yād kunad*

"Jāḥiẓ has mentioned **this report** in some of his books" (*Tārīḵnāme-ye Ṭabari*, ed. Rowšan 2001: IV/1098)

(5)

مر اندامش ایزد یکایک ستود

***mar andām-š** ēzad yakāyak sutūd*

"God extolled **his limbs** one by one" (*Garšāsb-Nāma*, ed. Yaġmāʾi 1938: 2)

(6)

مردی باشد بر ستوران تا **مر ایشان** از ددکان و درّندگان نگاه دارد

*mard-ē bāšad bar sutōrān tā **mar ēšān** az dadakān u darrandagān nigāh dārad*

"There is a man over the cattle to protect **them** from wild animals and predators" (*Vajh-e Din*, ed. Erāni 1924: 10)

(7)

بگوی یا محمد **مر جهودان را و ترسان را**

*bigōy yā Muhammad **mar juhūdān rā u tarsāʾān rā***

"Say, O Mohammad, **to the Jews and the Christians**" (*Tafsir-e Qorʾān-e Pāk*, ed. Minovi 1969: 104)

(8)

ساقیا **مر مرا** از آن می ده

*sāqiyā **mar ma-rā** az ān may dah*

"O wine-pourer! Give **me** from that wine!" (Abū Šakūr, in: Lazard 1982: II/80)

(9)

نزدیک آرند بهشت **مر پرهیزگاران را**

*nazdīk ārand bahišt **mar parhēzgārān rā***

"They will bring Paradise nigh **for the righteous**" (*Tarjome-ye Qorʾān-e Muze-ye Pārs*, ed. Ravāqi 1976: 96)

(10)

سبب بیرونی **مر گرمی را** سه گونه بود

*sabab-i bērōnī **mar garmī rā** si gōna buvad*

"The external causes **of heat** are of three types" (*Ṭabiʿiyyāt-e Dānešnāme-ye ʿAlāʾi*, ed. Meškāt 1952: 26)

(11)

ما **مر [این] دین را** مخالف نباشیم

*mā **mar [ēn] dēn rā** muxālif nabāšēm*

"We are not the opponents **of this religion**" (*Tāriḵ-e Balʿami*, ed. Bahār 1974: I/312)

(12)

ز ریگ ار فزون **مر شما را شمار**

*zi rēg ar fuzōn **mar šumā rā šumār***

"[even] if **your number** is more than the sand" (*Garšāsb-Nāma*, ed. Yaġmāʾi 1938: 88)

(13)

فلک محیط سطح **مر وی**

*falak-i muhīt-i sath-i **mar vay***

"The firmament surrounding **its** surface" (*Ḵān al-ʾIḵwān*, ed. Qavim 1959: 126)

(14)

من این کتاب **مر آنرا** ساختم که صقّال دلها بود

*man ēn kitāb **mar ān rā** sāxtam ki saqqāl-i dilhā buvad*

"I composed this book **for that [reason]** that it be a polisher of hearts" (*Kašf al-Maḥjūb*, ed. Žukovskij 1926: 5-6)

It is important to know that the particle *mar* is found abundantly in all Classical Persian texts from Transoxiana and the northeastern part of present-day Afghanistan, particularly in the works of Avicenna and Nāṣir-i Ḵusraw. It is much less attested in texts written in present-day Iran and the southern and western parts of present-day Afghanistan. Therefore, it seems proper to believe that *mar* had originally been a dialect peculiarity (cf. Lazard 1963: 382-383; Nātel-e Ḵānlari 1986: III/390; Maggi and Orsatti 2018: 41).

The particle *mar* is attested in some of the Judaeo-Persian translations of the Bible, e.g., in a manuscript preserved in the Bibliothèque nationale de France:

*wplyd krdyd **mr zmyn mn***
"and you defiled **my land**" (Jeremiah 2.7; Lagarde 1884: 64)

It is also attested in an Early New Persian manuscript in Syriac script discovered in Turfan. This manuscript is particularly valuable for the vocalization of the Persian words:

bzwᵘrg kwᵘnᵃδ xwδʾh ***mr drwyᵉšāʾn rāʾ***
"The Lord makes great **the poor**" (folio II, recto 3-4; Sims-Williams 2011: 357)

The particle *mar* might have also been used in the fragmented Early New Persian versification of the tale of *Bilawhar wa Būdīsaf* (Barlaam and Josaphat) in Manichaean script discovered in Turfan:

gw(f)[tyy] ***m[r mrʾ]***
"you said **to me**" (folio A, recto 2; Henning 1962: 94)

In addition to Early New Persian, the particle *mar* is attested once in a Sogdian text written in the city of Xumdān (i.e., 西安: Xī'ān) in China. Here *mar* seems to emphasize the following adverb:

rty ʾnyh tmyh ***mr zʾry*** *mrčh šw kʾmt rty Lʾ βyrt*
"And in another hell he yearns **pitifully** for his death, and does not get it" (folio XX, verso 1085-1086; Benveniste 1940: 51; cf. Gharib 1995: 215)

As it was said above, there is no trace of the Classical Persian particle *mar* in the extant Middle Persian texts, a fact that strengthens the probability of its being borrowed from a neighbouring language. Since the particle *mar* is abundant in all Classical Persian texts from Transoxiana and the northeastern part of present-day Afghanistan, the most appropriate candidate would be the Bactrian language, an Eastern Middle Iranian language spoken from about the 1st to the 9th century AD in a wide area in and around ancient Bactria in northern Afghanistan. The Bactrian language is the only Iranian language whose writing system is based on the Greek alphabet. It was one of the least-known Iranian languages until 1990's, when the unexpected discovery of a wealth of manuscripts in Afghanistan contributed significantly to our knowledge of this language. These manuscripts, written on leather, cloth, and even on wooden sticks, consist of legal documents, economic documents, letters, and Buddhist texts.

In my opinion, the Classical Persian particle *mar* is a focus marker[2] (cf. Lenepveu-Hotz 2018: 94-97) which ultimately goes back to the Bactrian locative adverb *μαρο [mar]*[3] "here", a well-attested word in Bactrian documents:

[2] Focus is an attention-getting mechanism which in spoken language is recognizable by, for example, putting stress on a word. It is, therefore, "dependent upon discourse structure but does not make up part of the structure itself" (Radetzky 2002: 103).

[3] In the Bactrian documents written in the Greek alphabet, virtually every word ends with a vowel letter, usually *-o* (Sims-Williams 2000: 24; Sims-Williams 2007: 40).

οτο αζο ***μαρο*** *λρογημο*
"and I am healthy **here**" (document bh 7; Sims-Williams 2007: 67)

αγαδο ***μαρο*** *ασο το χοηο πωστογο*
"a letter came **here** from your lordship" (document cd 3-4; Sims-Williams 2007: 75)

ταδο αβο χοβανανο ***μαρο*** *πισοαμαγο φοϸτιιο*
"then send the shepherds **here** into my presence" (document ba 13; Sims-Williams 2007: 53)

οτανο ***μαρο*** *αβο ρωβο αγαδινδηιο*
"and they have come **here** to (the city of) Rōb" (document cl 5; Sims-Williams 2007: 89)

Bactrian *μαρο [mar]* "here" goes back to Old Iranian **imaθra*[4], a combination of **ima* "this" and the suffix **-θra* which made locative adverbs[5]. The development of the Bactrian locative adverb *μαρο [mar]* into the Classical Persian focus marker *mar* can be easily compared with and explained by similar grammaticalization developments in other languages. The use of locative adverb as demonstrative, proximal or distal, is a well-known grammaticalization development attested in a number of languages from different language families (e.g., French, Hausa, Lingala, Ngbaka, Buang; see Heine and Kuteva 2002: 172-173, 294-295). On the other hand, the development from demonstrative to such grammatical items as definite article and focus marker is a common process in world languages (see, e.g., Diessel 1999: 155; Heine and Kuteva 2002: 109-112). It cannot, of course, be determined when exactly the development from locative adverb (*mar* "here") to proximal demonstrative (*mar* "this") and then to focus marker occurred in Classical Persian.

This final *-ο*, at least in some instances, has no phonetic value (Sims-Williams 1989: 348).

[4] For the phonological development of **θr* to *r*, the loss of initial **i-*, and the loss of final **-a* in Bactrian, see Gholami 2014: 52, 58, 61.

[5] Cf. the Sanskrit suffix *-tra* and the Avestan suffix *-θra* with the same function (Whitney 1879: 358; Jackson 1892: I/201). Old Iranian **imaθra* is also reflected in Khotanese *mara* "here" (Bailey 1979: 324; Sims-Williams 2000: 203; Sims-Williams 2007: 231) and Sogdian *mrδ [marθ]* (Gershevitch 1954: 67; Gharib 1995: 216). Bactrian *μαρο [mar]* "here" from Old Iranian **imaθra* is comparable with Bactrian *μαλο [mal]* "here" from Old Iranian **imada* (Sims-Williams 2000: 202; Sims-Williams 2007: 230; Gholami 2014: 58), the latter Old Iranian form is also reflected in Sogdian *[maδ]*, written in a variety of forms: *ʾmδ*, *ʾmδʾ*, *mδ*, *mδh*, *mδy*, *mdy*, *mδyy* (Gharib 1995: 34, 210, 211). There is no clear distinction between the function of *μαρο* and *μαλο* in Bactrian (Gholami 2014: 160).

It seems that the Parāchi preposition *ma*, like the Classical Persian focus marker *mar*, has developed similarly, but independently, from the Bactrian locative adverb *μαρο [mar]*. The Parāchi preposition *ma* is not only used in a local and temporal sense, but is also used before the specific direct object and the dative with "to give" and "to say": ***ma*** *dur* "**at** the river"; ***ma*** *dōwās* "**at** 12 o'clock", ***ma bālō*** *dhōṛ-um* "I saw **the boy**", ***ma puš-ē*** *ǰaṛī* "he said **to his son**" (cf. Morgenstierne 1929: 52; Morgenstierne 1985: 524; Efimov 1999b: 263; Kieffer 2009: 699). A similar grammaticalization development is seen in the Logar dialect of the Ormuṛi language, where the specific direct object is sometimes preceded by *ku*: *towa* ***ku-tsimi-m*** *roxšawok* "The sun blinded **my eyes**", ***ku kitāb*** *bu awīm* "I am reading **the book**" (Efimov 2011: 127; cf. Morgenstierne 1929: 343; Efimov 1999a: 284). This *ku*, like the Sogdian preposition *(ə)kū*[6] "to, towards", ultimately goes back to the Old Iranian locative adverb **ku* "where"[7] (see also Sims-Williams 1986: 118; Yoshida 2009: 293).

In conclusion, the Classical Persian particle *mar* is an optional focus marker which highlights the word it precedes. It ultimately goes back to the Bactrian locative adverb *μαρο [mar]* "here" which, as a result of grammaticalization, developed into a demonstrative and then a focus marker. Therefore, it has no etymological connection with its homonym *mar* "number; account", nor with the Early Judaeo-Persian preposition *azmar (i)* "for". Grammaticalization of locative adverbs is a common process in world languages. It can also be seen in the two Iranian languages of Parāchi and Ormuṛi, where a locative adverb has developed into a marker of the specific direct object. In Classical Persian the function of marking the specific direct object was already assigned to the postposition *rā*[8]; therefore, the Bactrian locative adverb *μαρο [mar]* which appeared as a loanword in the Persian dialects of Transoxiana and the northeastern part of present-day Afghanistan, assumed the function of a focus marker.

BIBLIOGRAPHY

Agostini, D. (2014), "Eschatological Seers and Otherworldly Travellers in Zoroastrianism", *Journal Asiatique*, vol. 302, pp. 47-73.

Bahār, M.-T. (1355/1976), *Sabkšenāsi*, 4th ed., 3 vols., Tehrān: Ketābhā-ye Parastu.

[6] Variously spelt as *ʾkʾw, ʾkw, kʾw, kw, kww, qw* (Gharib 1995: 31, 33, 187, 199, 202).

[7] Cf. Old Avestan *kū* "where?", *kudā* "whither?", *kuθrā* "where?", Young Avestan *kuua* "where?", *kuθra* "where?", *kudaţ* "wherefrom?" (see also Sanskrit *ku-* in Mayrhofer 1956: I/217).

[8] From Middle Persian *rāy* (MacKenzie 1986: 71), from Old Persian *rādiy/rādī* "for the sake of" (Kent 1953: 205; Schmitt 2014: 236).

Bailey, H. W. (1979), *Dictionary of Khotan Saka*, Cambridge: Cambridge University Press.

Beekes, R. (2010), *Etymological Dictionary of Greek*, 2 vols., Leiden: Brill.

Benveniste, É. (1938), "Sur un fragment d'un psautier syro-persan", *Journal asiatique*, vol. 230, pp. 458-462

Benveniste, É. (1940), *Textes sogdiens: édités, traduits et commentés*, Paris: Librairie Orientaliste Paul Geuthner.

Bossong, G. (1985), *Empirische Universalienforschung: Differentielle Objektmarkierung in den neuiranischen Sprachen*, Tübingen: Gunter Narr Verlag.

Cheung, J. (2007), *Etymological Dictionary of the Iranian Verb*, Leiden: Brill.

Dāʿī al-ʾIslām, M.-ʿA., *Farhang-e Neẓām*, 5 vols., Hyderabad (Deccan): Aʿẓam Steam Press/Šams al-Maṭābiʿ Machine Press, 1929-1932-1934-1938-1939.

Darmesteter, J. (1883), *Études iraniennes*, 2 vols., Paris: F. Vieweg.

Diessel, H. (1999), *Demonstratives: Form, Function, and Grammaticalization* (= Typological Studies in Language 42), Amsterdam: John Benjamins Publishing Company.

Efimov, V. A. (1999a), "Ormuri jazyk", in: *Jazyki mira: Iranskie jazyki II: Severo-zapadnye iranskie jazyki*, Moskva: Indrik, pp. 276-296.

Efimov, V. A. (1999b), "Parači jazyk", in: *Jazyki mira: Iranskie jazyki II: Severo-zapadnye iranskie jazyki*, Moskva: Indrik, pp. 257-275.

Efimov, V. A. (2011), *The Ormuri Language in Past and Present*, ed. J. L. G. Baart, Islamabad: Forum for Language Initiatives.

Enju Širāzi, M. J. Ḥ., *Farhang-e Jahāngiri*, ed. R. ʿAfifi, 3 vols., Mašhad: Dānešgāh-e Mašhad, 1351/1972.

Eṣfahāni, Ḥ. (1289/1872), *Dastur-e Soḵan*, Istānbul: Author.

Eṣfahāni, Ḥ. (1308/1890), *Dabestān-e Pārsi*, Istānbul: Čāpḵāne-ye Maḥmudbek.

Faršidvard, Ḵ. (1382/2003), *Dastur-e Mofaṣṣal-e Emruz bar Pāye-ye Zabān-Šenāsi-ye Jadid*, Tehrān: Entešārāt-e Soḵan.

Garšāsb-Nāma, ed. Ḥ. Yaḡmāʾi, Tehrān: Ketābforuši-ye Boruḵim, 1317/1938.

Gershevitch, I. (1954), *A Grammar of Manichean Sogdian*, Oxford: Basil Blackwell.

Gharib, B. (1995), *Sogdian Dictionary (Sogdian – Persian – English)*, Tehrān: Farhangān Publications.

Gholami, S. (2014), *Selected Features of Bactrian Grammar*, Wiesbaden: Harrassowitz Verlag.

Gignoux, Ph. (2010), "La collection de textes attribuables à Dādēn-vindād dans l'Archive pehlevie de Berkeley", in: *Sources for the History of Sasanian and Post-Sasanian Iran* (= Res Orientales XIX), Bures-sur-Yvette: Groupe pour l'Étude de la Civilisation du Moyen-Orient, pp. 11-134.

Gindin, Th. E. (2007), *The Early Judaeo-Persian Tafsīrs of Ezekiel: Text, Translation, Commentary*, 3 vols., Wien: Österreichische Akademie der Wissenschaften.

Gray, L. H. (1937), "Notes on Modern Persian Morphology", *Journal of the Royal Asiatic Society*, vol. 69, part II, pp. 305-309.

Heine, B., and Kuteva, T. (2002), *World Lexicon of Grammaticalization*, Cambridge: Cambridge University Press.

Henning, W. B. (1962), "Persian Poetical Manuscripts from the Time of Rūdakī", *A Locust's Leg, Studies in Honour of S. H. Taqizadeh*, London: Percy Lund, Humphries & Co. Ltd., 89-104.

Homāyunfarroḵ, ʿA. (1339/1960), *Dastur-e Jāme-ʾe Zabān-e Fārsi*, 2nd ed., Tehrān: Entešārāt-e ʿElmi.

Horn, P. (1893), *Grundriss der neupersischen Etymologie*, Strassburg: Karl J. Trübner.

Horn, P. (1898), "Neupersische Schriftsprache", in: *Grundriss der iranischen Philologie*, Band I, 2. Abteilung, ed. W. Geiger and E. Kuhn, Strassburg: Karl J. Trübner, pp. 1-200.

Jackson, A. V. W. (1892), *An Avesta Grammar in Comparison with Sanskrit (Part I: Phonology, Inflection, Word-Formation)*, Stuttgart: Kohlhammer.

Jāmiʿ al-Ḥikmatayn, ed. H. Corbin and M. Moʿin, Tehrān: Institut Français de Recherche en Iran, 1332/1953.

Jones, W. (1771), *A Grammar of the Persian Language*, London: W. and J. Richardson.

Ḵān al-ʾIḵwān, ed. ʿA. Qavim, Tehrān: Ketābḵāne-ye Bārāni, 1338/1959.

Kašf al-Maḥjūb, ed. V. A. Žukovskij, Leningrad: Gosudarstvennaja akademičeskaja tipografija, 1926.

Kent, R. G. (1953), *Old Persian Grammar, Texts, Lexicon*, 2nd ed., New Haven: American Oriental Society.

Ḵatib-e Rahbar, Ḵ. (1367/1988), *Dastur-e Zabān-e Fārsi: Ketāb-e Ḥoruf-e Ezāfe va Rabṭ*, Tehrān: Entešārāt-e Saʿdi.

Kieffer, Ch. M. (2009), "Parachi", in: *The Iranian Languages*, ed. G. Windfuhr, London: Routledge, pp. 693-720.

Lagarde, P. de (1884), *Persische Studien*, Göttingen: Abhandlungen der historisch-philologischen Classe der königlichen Gesellschaft der Wissenschaften zu Göttingen XXXI.2.

Lazard, G. (1963), *La langue des plus anciens monuments de la prose persane*, Paris: Librairie C. Klincksieck.

Lazard, G. (1982), *Les premiers poètes persans (IXe-Xe siècles),* 2nd ed., 2 vols., Tehrān: Institut Français de Recherche en Iran.

Lazard, G. (2009), "Qu'est devenue la préposition *ō?",* in: *Zarathushtra entre l'Inde et l'Iran: Études indo-iraniennes et indo-européennes offertes à Jean Kellens à l'occasion de son 65e anniversaire,* ed. É. Pirart and X. Tremblay, Wiesbaden: Dr. Ludwig Reichert Verlag, pp. 169-176.

Lenepveu-Hotz, A. (2018), "Specialization of an Ancient Object Marker in the New Persian of the Fifteenth Century", in: *Trends in Iranian and Persian Linguistics* (= Trends in Linguistics: Studies and Monographs, vol. 313), ed. A. Korangy and C. Miller, Berlin: Walter de Gruyter, pp. 81-100.

MacKenzie, D. N. (1968), "An Early Jewish-Persian Argument", *Bulletin of the School of Oriental and African Studies*, vol. 31, no. 2, pp. 249-269.

MacKenzie, D. N. (1986), *A Concise Pahlavi Dictionary*, 2nd ed., London: Oxford University Press.

Macuch, M. (2008), "Anhang VI: Zur *juristischen* Terminologie der Berliner Pahlavi-Dokumente", in: *Berliner Pahlavi-Dokumente: Zeugnisse spätsassanidischer Brief- und Rechtskultur aus frühislamischer Zeit* (= Iranica 15), Wiesbaden: Harrassowitz Verlag, pp. 249-266.

Maggi, M., and Orsatti, P. (2018), "From Old to New Persian", in: *The Oxford Handbook of Persian Linguistics*, ed. A. Sedighi and P. Shabani-Jadidi, Oxford: Oxford University Press, pp. 7-51.

Maškur, M.-J. (1340/1961), *Dastur-Nāme dar Ṣarf va Naḥv-e Zabān-e Pārsi*, 2nd ed., Tehrān: Moʾassese-ye Maṭbuʾāti-ye Šarq.

Mayrhofer, M. (1956-1976), *Kurzgefaßtes etymologisches Wörterbuch des Altindischen*, 3 vols., Heidelberg: Carl Winter.

Moʾayyedi, M., and Loṭfi, ʾA.-R. (1392/2013), "Barrasi-ye Sāḵt-e Do-Mafʿuli dar Motun-e Adab-e Fārsi", *Pažuhešhā-ye Zabān-Šenāsi*, year 5, no. 1, pp. 101-119.

Moḥammad-Ḥosayn b. Ḵalaf Tabrizi, *Borhān-e Qāteʿ*, ed. M. Moʿin, 2nd ed., 4 vols., Tehrān: Ketābforuši-ye Ebn-e Sinā, 1342/1963.

Morgenstierne, G. (1929-1967), *Indo-Iranian Frontier Languages, vol. I: Parachi and Ormuri* (1929), *vol. II: Iranian Pamir Languages (Yidgha-Munji, Sanglechi-Ishkashmi and Wakhi)* (1938), *vol. III: The Pashai Language, 2. Texts and Translations* (1944), Oslo: H. Aschehoug & Co. (W. Nygaard); *vol. III: The Pashai Language, 1. Grammar* (1967), Oslo: Universitetsforlaget.

Morgenstierne, G. (1985), "Parāčī", *Encyclopaedia Iranica*, ed. E. Yarshater, vol. I, London/New York: Routledge & Kegan Paul, pp. 522-525.

Nafisi, ʿA.-ʾA. (1355/1976), *Farhang-e Nafisi*, 5 vols., Tehrān: Ketābforuši-ye Ḵayyām.

Nātel-e Ḵānlari, P. (1365/1986), *Tāriḵ-e Zabān-e Fārsi*, revised ed., 3 vols., Tehrān: Našr-e Now.

Parizāde, M. (1399/2020), "Dasturi-Šodegi-ye *mar* va Pirā-Ezāfe-ye *mar … rā* dar Zabān-e Fārsi", *Pažuhešhā-ye Zabān-Šenāsi-ye Taṭbiqi*, year 10, no. 20, pp. 221-242.

Paul, L. (2013), *A Grammar of Early Judaeo-Persian*, Wiesbaden: Reichert Verlag.

Phillott, D. C. (1919), *Higher Persian Grammar*, Calcutta: The University.

Platts, J. T. (1894), *A Grammar of the Persian Language*, London: Williams and Norgate.

Radetzky, P. K. (2002), *The Functions and Evolution of Topic and Focus Markers*, Unpublished PhD Dissertation in Linguistics, University of California, Berkeley.

Rückert, Fr. (1854), "Bemerkungen zu Mohl's Ausgabe des Firdusi, Band I", *Zeitschrift der Deutschen Morgenländischen Gesellschaft*, vol. 8, no. 2, pp. 239-329.

Ṣāḥebi, M. (1397/2018), "Vižegihā-ye Dasturi-ye Kamyāb va Barjaste dar Tarjome-ye Adʿiye-ye Ḏaḵīrat al-ʾĀḵirah", *Dastur*, no. 14, pp. 17-49.

Salemann, C. (1895), "Mittelpersisch", in: *Grundriss der iranischen Philologie*, Band I, 1. Abteilung, ed. W. Geiger and E. Kuhn, Strassburg: Karl J. Trübner, pp. 249-332.

Schmitt, R. (2014), *Wörterbuch der altpersischen Königsinschriften*, Wiesbaden: Reichert Verlag.

Shaked, Sh. (2003), "Early Judaeo-Persian Texts with Notes on a Commentary to Genesis", in: *Persian Origins: Early Judaeo-Persian and the Emergence of New Persian*, ed. L. Paul, Wiesbaden: Harrassowitz Verlag, pp. 195-219.

Sims-Williams, N. (1986), "Sogdian *kw* and Slavonic *kū*", *Peredneaziatskij sbornik* IV, pp. 116-121.

Sims-Williams, N. (1989), "Bactrian Language", *Encyclopaedia Iranica*, ed. E. Yarshater, vol. III, London and New York: Routledge & Kegan Paul, pp. 344-349.

Sims-Williams, N. (2000), *Bactrian Documents from Northern Afghanistan, I: Legal and Economic Documents* (= Studies in the Khalili Collection, vol. III, Corpus Inscriptionum Iranicarum, part II, vol. VI), Oxford: Oxford University Press.

Sims-Williams, N. (2007), *Bactrian Documents from Northern Afghanistan, II: Letters and Buddhist Texts* (= Studies in the Khalili Collection, vol. III, Corpus Inscriptionum Iranicarum, part II, vol. III), London: The Nour Foundation.

Sims-Williams, N. (2011), "Early New Persian in Syriac Script: Two Texts from Turfan", *Bulletin of the School of Oriental and African Studies*, vol. 74, no. 3, pp. 353-374.

Steingass, F. J. (1892), *A Comprehensive Persian-English Dictionary*, London: Routledge & Kegan Paul Limited.

Šafiʿi, M. (1343/1964), *Šāhnāme va Dastur*, Tehrān: Entešārāt-e Nil.

Ṭabiʿiyyāt-e Dānešnāme-ye ʿAlāʾi, ed. M. Meškāt, Tehrān: Anjoman-e Āṯār-e Melli, 1331/1952.

Tafsir-e Qorʾān-e Pāk, ed. M. Minovi, Tehrān: Bonyād-e Farhang-e Irān, 1348/1969.

Tāriḵ-e Balʿami, ed. M.-T. Bahār, 2nd ed., 2 vols., Tehrān: Ketābforuši-ye Zavvār, 1353/1974.

Tāriḵnāme-ye Ṭabari, ed. M. Rowšan, 3rd ed., 5 vols., Tehrān: Entešārāt-e Soruš, 1380/2001.

Tarjome-ye Qorʾān-e Muze-ye Pārs, ed. A. Ravāqi, Tehrān: Bonyād-e Farhang-e Irān, 1355/1976.

Tarjome-ye Tafsir-e Ṭabari, ed. H. Yaḡmāʾi, 2nd ed., 7 vols., Tehrān: Entešārāt-e Tus, 1356/1977.

Vahman, F. (1986), *Ardā Wirāz Nāmag: The Iranian Divina Commedia*, London: Curzon Press.

Vajh-e Din, ed. T. Erāni, Berlin: Kaviani GmbH, 1343/1924.

Weber, D. (2008), *Berliner Pahlavi-Dokumente: Zeugnisse spätsassanidischer Brief- und Rechtskultur aus frühislamischer Zeit* (= Iranica 15), Wiesbaden: Harrassowitz Verlag.

Whitney, W. D. (1879), *A Sanskrit Grammar*, Leipzig, Breitkopf and Härtel.

Yoshida, Y. (2009), "Sogdian", in: *The Iranian Languages*, ed. G. Windfuhr, London: Routledge, pp. 279-335.

YEREVAN STATE UNIVERSITY PUBLISHING HOUSE

JIL 1 (2024): 35-65

Journal of Iranian Linguistics

On the Syntax of the Persian Classical Narrative Poetry: Constructions with a Past Participle in the *Shāhnāme*

Paola Orsatti

Sapienza University of Rome

doi.org/ 10.46991/jil/2024.01.03

Abstract. This paper aims to describe a little studied syntactic construction of Early and Classical New Persian which involves dependent phrases or clauses with a verb in the form of a past participle. The examples are mainly taken from Ferdowsi's *Shāhnāme* and texts of early poetry, where the construction is well-attested, but examples from early prose texts have also been given. This shows that the construction was not restricted to poetry. In the *Shāhnāme* and early texts, participial constructions are usually placed after a clause with a finite verb in the past tense. They are endowed with a series of syntactic and semantic functions, such as that of adnominal or adverbial modification through phrases or clauses, or predicative complement. Of these, only constructions with an adverbial value, mainly expressing time, but also manner, cause, and other relations, continue up to the present, though in a more rigid form: they are placed before the main clause, and have the value of a subordinate adverbial clause indicating anteriority to the action of the main clause (e.g. *be khāne* ***rafte***, *shām khwordam* 'when I went home, I had dinner'). At the end of the article, a hypothesis aimed at explaining the development of the construction from Early and Classical to Modern New Persian is put forward.

Keywords: *Shāhnāme*; Early and Classical New Persian; participial constructions; past participle; absolute constructions; predicative usages; adverbial clauses

Paola Orsatti

E-mail:
paola.orsatti@uniroma1.it

Received: 06.03.2024
Revised: 02.07.2024
Accepted: 11.07.2024

This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License

Introduction

To Djalal Khaleghi Motlagh
in deep admiration and gratitude
for his invaluable contribution
to Persian studies

1. Introduction[1]

Narrative poetry often shows a syntactic complexity unknown to lyric poetry. Along with simple syntactic constructions, in which each line, or each half-line, represents a syntactic unit, in the *Shāhnāme* by Ferdowsi (composed between the last decades of the tenth and the early eleventh century) more complex passages stretching over two or more lines are found. These include dependent clauses and parenthetical expressions whose structure is not always easy to understand. We do not know how written texts of early epochs were read or recited. Thus, linguistic features such as intonation and pauses, truly important for a correct understanding of the syntactic structure and the meaning of the text, have been lost.

This paper aims to analyse dependent syntactic constructions with a past participle (past stem + *-e*) in texts of New Persian early poetry.[2] This topic has received only limited attention by scholars (see below). For pre-modern stages of New Persian, these constructions pose a problem of identification and linguistic interpretation. The present study has a mainly descriptive character: participial constructions are analysed through examples from the *Shāhnāme* and other early texts. The question of the origin of the Early and Classical New Persian participial constructions, a type of construction widely

[1] I would like to thank Agnès Lenepveu-Hotz, who read a first draft of this article, and Giacomo Brotto, who supplied a couple of examples from the *Shāhnāme*. The transcription of Early and Classical texts is given according to the modern pronunciation of New Persian and, for poetry texts, reflects the metrical spelling of the lines; e.g. *sov=i* (as a sequence of a short and a long syllable, in example [1]), instead of *su=yi* 'side=ART.INDF'. To contextualize the examples from the *Shāhnāme* it is often necessary to provide narrative context.

[2] In New Persian (henceforth, if not otherwise specified, simply Persian) the past participle is formed from the past stem of verbs (Middle Persian past participle) followed by the suffix *-e* (Middle Persian adjectival suffix *-ag*). Persian has no nominal inflection (apart from plural marking) and no distinction of grammatical gender. Past participles, as all adjectives, do not agree in number. Past participles from transitive verbs can have an active, intransitive/stative, or passive value according to their syntactic function, and sometimes according to the semantics of each individual verb (see below, fn. 25). An analysis of constructions with present participles, which have a more limited usage, has not been provided in this study.

attested in ancient Indo-European languages (see, among others, Holland 1986), is left for future research.

Constructions with a verb in the past participle are a type of syntactic construction known to Modern New Persian, though perceived as literary or obsolete today: *be khāne* ***rafte***, *shām khordam* 'having gone home/ when I went home/ after I went home, I had dinner'. In this kind of construction, the syntactic unit with a verb in the past participle is a subordinate adverbial clause mainly expressing a circumstance of time. It precedes the main clause and indicates anteriority to the action of the main clause. A comparison with participle clauses with a perfect participle in English (e.g. *Having won the match, Susan jumped for joy*) may be useful. With set expressions, the dependent adverbial clause may have a different subject from that of the main clause, as in *do sāʿat be ghorub mānde be Eṣfahān residim* 'we arrived in Isfahan two hours before sunset (lit. two hours having remained to sunset)'.

As will be shown below, in Early and Classical New Persian texts dependent constructions with a past participle are much more frequent than in the contemporary language, and are endowed with a wider range of meanings and functions. Unlike the modern language, they normally follow a clause with a finite verb, which is generally in the past tense. Moreover, in Early and Classical New Persian texts, participial constructions with a subject differing from that of the superordinate clause are of normal usage.

For the modern language, this kind of construction has been especially described by Hans Jensen (1931: 250-251), Gilbert Lazard (2006: 155-156, 189-191, 204), Jurij A. Rubinčik (2001: 276). Suggestions on historical development are to be found in the works by Parviz Nātel Khānlari (1986: vol. 3, pp. 456-457 §§ 9.5 and 9.6), Khosrow Farshidvard (1999: 324-327, 401-408), Ḥasan Aḥmadi Givi (2001: vol. 1, pp. 728-761), and in Alessandro Bausani's recently published thesis (tesi di laurea) from 1943 featuring a historical linguistic perspective (see Bausani 2023: 83, 85-86, 92-93 [=39, 41-42, 48-49]). Brief remarks on the use of participial expressions in Early and Classical New Persian are also given by Moḥammad-Taqi Bahār (1958: vol. 2, pp. 76, 256, and *passim*), and by Gernot Windfuhr (1979: 75). Agnès Lenepveu-Hotz, in her important work on the history of the New Persian verbal system, only hints at them (2014: 168 example 4). The usage of past participles in these constructions is not described by Gilbert Lazard (1963) in his invaluable description of the linguistic characters of literary Early New Persian prose texts. For a general discussion of the studies on participial constructions see Orsatti (2023: 114-121 [=12-19]).

The participle in participial constructions should not be confused with the participle of perfect forms with an implicit or dropped auxiliary, either coordinated with a form with auxiliary, or, mainly in the 3rd person singular,

used alone (for the latter usage cf. Lenepveu-Hotz 2014: 167-168). The omission of an auxiliary in coordinated compound verbal forms falls within the important stylistic character of Persian prose referred to as *ḥadhf-e afʿāl be qarine[-ye afʿāl-e digar]* 'omission of verbs on the basis of the context' (see Bahār 1958: vol. 2, pp. 73-74 and *passim*; Khānlari 1986: vol. 2, p. 392; vol. 3, p. 471). For example, in *sharāb dar-u asar* ***karde bud*** *va ʿeshq dar-u ʿamal* ***nemude*** 'the wine had begun to affect him and love to stir within him'[3] the participle *nemude* is understood to be accompanied by the implicit auxiliary *bud* 'was' in the pluperfect tense.[4] In contrast, in the syntactic constructions analysed here the past participle may stand as the verb of a nominalized relative clause (with no relative pronoun and no auxiliary), or, more often, as a converb.[5] In reference to the converb-like function of past participles, the terms *feʿl-e vaṣfi* 'descriptive verb', *vajh-e vaṣfi* 'descriptive mood', or *sighe-ye vaṣfi* 'descriptive form' are used in the Persian grammatical tradition.[6]

The frequent occurrence of participial constructions in narrative poetry, especially in descriptive passages that unfold over several lines, responds to the need of providing the reader or hearer with a lively yet compact description. As will be shown, similar constructions are attested in coeval prose texts too. This seems to indicate that a construction that existed in the common language, but was possibly somewhat rare, was exploited by the poets, in narrative poems, as a mainly stylistic tool.

2. Examples from Ferdowsi's Shāhnāme

Syntactic constructions (clauses, phrases) with a verb in the form of a past participle are frequent in the *Shāhnāme*.[7] As already stated, they generally

[3] Neẓāmi ʿArużi, *Chahār maqāle*, ed. Moḥammad Qazvini: 34; transl. by Edward G. Browne: 56.

[4] In Persian the auxiliary of perfect forms is *budan* 'to be'.

[5] For a definition of 'converb', i.e. verbal adverb, a non-finite verb form that expresses adverbial subordination, see Haspelmath 1995. On the reasons for the choice of such a neologism in reference to a form variously referred to as 'gerund' in Italian and some Romance languages, 'gérondif' in French, 'adverbial participle' in some studies, etc., see Haspelmath 1995: 45-46 ('gerund' in English denotes a verbal noun, as in '*Eating* ice cream on a hot day is refreshing'). In Persian, the term *feʿl-e vaṣfi* is used (see the following footnote).

[6] Farshidvard (1999: 401) considers the past participle in this function as a grammatical mood, besides the indicative, subjunctive, imperative, and infinitive, and explains: "The *vajh-e vaṣfi* or *feʿl e vaṣfi* is a past participle (*esm-e mafʿul*) that plays the role of a verb". I translate *vaṣfi* as 'descriptive', though *vaṣfi* can also be translated as 'qualificative', or 'adjectival'.

[7] Single lines with past participles in the function of *feʿl-e vaṣfi* from the *Shāhnāme* are quoted by Shafiʿi (1964: 234-235) and by Farshidvard (1999: 324-326).

follow the clause with a verb of finite form in the past. As for their function, they seem comparable to the modern adverbial clauses with a past participle in the function of a converb (see above, fn. 5), or, more rarely, to a relative clause with an implicit relative pronoun and an implicit auxiliary verb. Very often, past participles in the *Shāhnāme* are part of predicative expressions.

In the following passage, after the *farrah* ('divine charisma') of Jamshid became dark, Ferdowsi says:[8]

[1]

170 پدید آمد از هر سوی خسروی یکی نامجویی به هر پهلوی

171 سپه **کرده** و جنگ را **ساخته** دل از مهر جمشید **پرداخته**.[9]

170A	*padid*	*āmad*	*az*	*har*	*sov=i*
	in.sight	come.PST[3SG]	from	every	side=ART.INDF
	khosrov=i				
	king[SBJ]=ART.INDF				

From all sides a king came forward,

170B	*yek=i*	*nāmjuy=i*	*be*	*har*
	one=ART.INDF	intrepid.man=ART.INDF	to	every
	pahlov=i			
	side=ART.INDF			

– an intrepid man on every side –

171A	*sepah*	***kard-e***	*vo*	*jang=rā*	***sākht-e***
	army[OBJ]	do.PST-PTCP	and	war=PURP	prepare.PST-PTCP

[who had] **raised** an army and [was] **ready** for war

171B	*del*	*az*	*mehr=e*	*Jamshid*	***pardākht-e***
	heart[OBJ]	from	love=of	Jamshid	free.PST-PTCP

[who had] **freed** his heart from loyalty to Jamshid.

In line 171, the past participles *karde* and *pardākhte*, from transitive verbs, have an active meaning, and indicate an action accomplished in an earlier time by the same subject as that of the verb of finite form (a king came forward, line 170A). They can be explained and rephrased as pluperfect forms (*karde [bud]*, *pardākhte [bud]*) of relative clauses modifying the subject in the

[8] For the *eżāfe* particle in the function of linking a substantive to its adjectival determinant, the label EZF has been introduced in the glosses.

[9] Ferdowsi, *The Shahnameh*, vol. 1, pp. 51 (*Jamshid*, lines 170–1).

finite clause: from everywhere a king, '**who had** raised (*karde*) an army' and '**who had** freed (*pardākhte*) his heart from loyalty', came forward (*padid āmad* 'lit. appeared').[10] But the two participles in the above example can also function as adverbial modifiers of the predicate in the finite clause. They refer to a circumstance of time related to the action expressed by the latter: from everywhere a pretender to the throne of Iran came forward, '**after** having raised an army' and '**after** having abandoned the covenant of loyalty to Jamshid'. Indeed, a distinction between the adnominal and adverbial functions of a past participle is not always clear-cut.

In contrast, in the combination *jang-rā sākhte* (171A) the past participle *sākhte* 'prepared, ready, disposed to', also from a transitive (and reflexive) verb, expresses a state, possibly resulting from a past action (the king **had prepared** for war, and therefore **was ready**), and has an intransitive value.[11] It is employed as an adjective taking a complement, that predicatively ascribes a quality or a manner of being to the subject of the finite verb form: from everywhere a pretender to the throne came forward, [who was/ being] ready for war.

In line 171 the participles used in what appears to be two distinct functions are coordinated, as if their different grammatical functions (as a verbal transitive active form, or as a lexicalized adjective denoting a state or a quality) were felt to be equivalent. Indeed, all three expressions with a participle show a greater or lesser degree of nominalization, and have a more general predicative function: they complete the meaning of the finite verb (*padid āmad* '[a king] appeared, came forward'), and 'predicate' a series of qualities and states referred to, or referentially controlled by, the subject of the finite verb (see also example [2] below and fn. 14).

In example [1] one could be led to suppose that the adnominal relativizing function possibly ascribed to the participles *karde* and *pardākhte* depends on the word order in line 170, with the subject, 'a king', immediately followed (after the parenthetical and elliptic sentence 'an intrepid man all around'), by the participial constructions, and functioning as the head noun of a relative clause.[12] The following example (example [2]), however, shows that the adnominal or adverbial function that can be ascribed to a past participle depends on an interplay of syntactic as well as semantic-contextual factors. In example [2] Khosrow Parviz entrusts a letter to an

[10] On participial relative clauses see Shagal 2019, in particular pp. 21-30. The objects governed by the two participles, *sepah* 'army' and *del* 'heart' respectively, are generic and therefore grammatically unmarked.

[11] The past participle *sākhte* is lexicalized as an adjective meaning 'ready' (cf. Anvari, *Farhang*, vol. 5, pp. 3965-3966: "5. *mohayyā, āmāde* 'ready, prepared'").

[12] In Standard New Persian the normal word order is considered to be SOV, with, however, a remarkable freedom, especially in Early New Persian prose texts and, of course, in poetry.

envoy (*navand*), to be delivered to his general-in-chief. But the envoy is caught and brought in front of the Qeyṣar, the Byzantine emperor:

[2]

3878 برون آمد از پیش خسرو نوند به بازو مر آن نامه را **کرده بند**

3879 بیامد چو نزدیک قیصر رسید یکی کارجویش به ره بر بدید

3880 سوی قیصرش برد سر پر ز گرد دو رخ زرد و لبها **شده لاژورد**.[13]

3878A	*berun*	*āmad*	*az*	*pish=e*	*Khosrow*
	out	come.PST[3SG]	from	in.front=of	Khosrow

navand
envoy[SBJ]

The envoy took his leave from Khosrow

3878B	*be*	*bāzu*	*mar*	*ān*	*nāme=rā*	***kard-e***
	to	arm	OBJ	that	letter=OBJ	do.PST-PTCP

band
bond

[after having] **tied** that letter to his arm.

3879A	*biy-āmad*	*cho*	*nazdik=e*	*Qeyṣar*
	out-come.PST[3SG]	when	near=of	the.Qeyṣar

rasid
arrive.PST[3SG]

He left. When he arrived in the vicinity of the Byzantine emperor,

3879B	*yek=i*	*kārjuy=ash*	*be*	*rah*
	one[SBJ]=ART.INDF	agent[SBJ]=him[OBJ]	in	road

bar	*be-did*
on	PFV-see.PST[3SG]

one agent [of the Byzantine emperor] saw him on the way.

3880A	*su=ye*	*Qeyṣar=ash*	*bord*	*sar*
	direction=of	the.Qeyṣar=him[OBJ]	take.PST[3SG]	head

por	*ze*	*gard*
full	of	dust

[The agent] took him to the Byzantine emperor, [his] head covered in dust,

[13] Ferdowsi, *The Shahnameh*, vol. 8, pp. 301-302 (*Khosrow Parviz*, lines 3878–3880).

3880B	*do*	*rokh*	*zard*	*o*	*lab-hā*	***shod-e***
	two	cheek	yellow	and	lip-PL	become.PST-PTCP
	lāzhvard					
	blue					

[his] cheeks yellow, and [his] lips [having] **turned** blue.

The participial construction in line 3878B (*be bāzu mar ān name-rā karde band* 'tied that letter to [his] arm') immediately follows the subject (*navand* 'the envoy', 3878A) of the finite clause, which is postponed to the finite verb (*berun āmad* 'took leave'), as in example [1]. However, the semantic context suggests that an adverbial (temporal) rather than adnominal (relative) function should be attributed to the participle: the envoy took leave from the king **after** having tied the letter to his harm. The participle, from a transitive verb (*band kardan* 'to tie'), governs a specific, and therefore marked, direct object: *mar ān nāme-rā* 'that letter'. The subject of the participle is Khosrow's envoy, as in the main clause. The participle is a *participium conjunctum*, as opposed to a *participium absolutum*, i.e. a participle having its own subject, different from the subject of the main clause (see below for examples).

In line 3880 Khosrow's envoy is described by a series of predicative expressions indicating a state or a manner of being of the person referred to. The envoy, the accusative experiencer in the finite clause (*-ash*, 3880A), was taken to the Qaysar:

1. *sar por ze gard*	[having] a head covered in dust,
2. *do rokh zard*	[being] yellow on cheeks/ [with] yellow cheeks,
3. *labhā* ***shode*** *lāzhvard*	[his] lips [having] **turned** blue.

The three noun phrases above are formally free from any bond with the rest of the sentence. The participle in 3, from a linking verb (*shodan* 'to become'), modifies the envoy's lips that had become, or, as an actual state resulting from a past action, **were** blue. If it were not for the meter, *shode* 'become' could have been omitted, without the meaning of the phrase being altered. Conversely, the past participle *shode* can be implied in the first two phrases as well: 1. *sar por ze gard* ***shode*** 'his head [having] **become** [i.e. being] covered in dust', and 2. *do rokh zard* ***shode*** 'his cheeks [having] **become** yellow', each having a subject differing from that of the verb of the finite clause.

The three predicative expressions above share with the converb-like construction in line 3878 a common general function. They 'predicate' a manner of acting or being of the envoy, who is the subject (the envoy took leave of Khosrow) in the first construction, and the object in line 3880 ([the agent] took **him** [i.e. the envoy] to the Qeyṣar). The finite verbs *berun āmad* '[he] took leave' and *bord* '[he] took' are not linking verbs. Even though they retain their ordinary full meaning, they can be modified by a predicative or *copredicative* expression. This is an expression that, together with the predicate, describes the manner an action is accomplished by the subject (or object).[14] The converb-like use of participles (as in 3878) is not too different, then, from the copredicative use of adjectives and adjective phrases, and of nouns and noun phrases (as in 3880), as the above example shows.[15]

The different value of a past participle as a predicative adjective (example [3]) and as a converb (example [4]) is illustrated by the following two examples, taken from the saga of the hero Rostam. When Rostam prepares a bed for himself and gets ready to sleep in a reed thicket infested with ferocious lions, a lion approaches and sees him asleep:

[3]

برِ نی یکی پیلتن **خفته** دید به پیشش یکی شیر **آشفته** دید.[16]

bar=e	*ney*	*yek=i*	*pil-tan*
side=of	reed	one[OBJ]=ART.INDF	elephant-bodied[OBJ]

khoft-e	*dīd*
fall.asleep.PST-PTCP	see.PST[3SG]

Next to the reeds, [the lion] saw one with an elephant's body [i.e. Rostam] **asleep**,

[14] All adjectives, including participles in their basic function as verbal adjectives, can be used predicatively (on the predicative, or copredicative use of participles, see Haspelmath 1995: 17-20). Examples of copredicative adjectives are the following: *Zhangsan came home* ***drunk***, and *Shanti drinks the milk* ***warm*** **(**Haspelmath 1995:18). Cf. also the example *She returned* ***a full-grown woman*** given by Jespersen (1933: 124 §13.2.1), who terms such expressions as 'quasi-predicatives' in as much as they can be rephrased by means of 'to be' and a predicative.

[15] As underlined by Haspelmath (1995), copredicative adjective phrases and noun phrases share several common features with converbal constructions, the most important of which – for the present study – is that "syntactically they depend on the predicate rather than on the controller of their implicit subject" (pp. 18-19).

[16] Ferdowsi, *The Shahnameh*, vol. 2, p. 22 (*Key Kāus*, line 289).

be	*pish=ash*	*yek=i*	*shir*	***āshoft-e***
in	front.of=him	one[OBJ]=ART.INDF	lion[OBJ]	agitate.PST-PTCP

did
see.PST[3SG]

in front of him he [the lion] saw a lion [Rostam] **restless** [in his sleep].

In this example, the past participles *khofte* 'asleep' from *khoftan* 'to sleep, to fall asleep', and *āshofte* 'troubled, agitated' from *āshoftan* 'to agitate; to disturb, be disturbed', are adjectives in a predicative function. They are co-referential with the object, Rostam, of the verb *didan* 'to see' and explain the way Rostam appeared to the anthropomorphized lion.

When Rostam finds the White Demon hidden in a pit:

[4]

به تاریکی اندر یکی کوه دید سراسر **شده** چاه ازو ناپدید.[17]

be	*tāriki*	*andar*	*yek=i*	*kuh*	*did*
in	darkness	inside	one[OBJ]=ART.INDF	mountain[OBJ]	see.PST[3SG]

In the darkness he [Rostam] saw a mountain [i.e. the Demon]

sar-ā-sar	***shod=e***	*chāh*	*az*	*u*	*nā-padid*
end-to-end	become.PST-PTCP	pit[SBJ]	from	him	NEG-in.sight

the pit [having] **become** entirely indistinguishable from him [the Demon].

The past participle (*nāpadid*) *shode*, from *nāpadid shodan* 'to be, become invisible, disappear', functions as the converb of an absolute construction with 'the pit' as subject.

The following example, from the description of the beginning of Jamshid's reign, offers some clear examples of absolute participial constructions, i.e. constructions each having their own subject, differing from the subject of the verb of finite form:

[5]

4 بر آمد بر آن تختِ فرخ پدر به رسم کیان بر سرش تاج زر
5 کمر **بسته** با فرّ شاهنشهی جهان **گشته** سرتاسر او را رهی
6 زمانه **بر آسوده** از داوری به فرمان او دیو و مرغ و پری

[17] Ferdowsi, *The Shahnameh*, vol. 2, p. 42 (*Key Kāus*, line 568).

7 جهان را **فزوده** بدو آبروی فروزان **شده** تخت شاهی بدوی.[18]

4A	*bar*	*āmad*	*bar*	*ān*	*takht=e*	*farrokh*	*pedar*
	up	come.PST[3SG]	on	that	throne=of	glorious	father

He [Jamshid] ascended the throne of his glorious father [Tahmurat],

4B	*be*	*rasm=e*	*key-ān*	*bar*	*sar=ash*	*tāj=e*	*zar*
	in	rule=of	king-PL	on	head=POSS.3SG	crown[SBJ]=of	gold

according to the rule of kings [with] a golden crown on his head,

5A	*kamar*	***bast-e***	*bā*	*farr=e*	*shāhanshahi*
	belt[OBJ]	fasten.PST-PTCP	with	glory=EZF	royal

[having] got ready for action [lit. having **fastened** the belt] with royal glory,

5B	*jahān*	***gasht-e***	*sar-tā-sar*	*u=rā*	*rahi*
	world[SBJ]	become.PST-PTCP	end-to-end	him=BEN	servant

the world [having] **become** his servant from end to end,

6A	*zamāne*	*bar*	***āsud-e***	*az*	*dāvari*
	time[SBJ]	up	rest.PST-PTCP	from	contention

time [having] **found rest** from [all] contention,

6B	*be*	*farmān=e*	*u*	*div*	*o*	*morgh*	*o*
	at	order=of	him	demon[SBJ]	and	winged.being[SBJ]	and

pari
fairy[SBJ]

[being] at his orders the demons, the winged beings and the fairies,

7A	*jahān=rā*	***fozud-e***	*bed-u*	*ābruy*
	world=BEN	increase.PST-PTCP	thanks.to-him	honour[SBJ]

the honour of the world [having] **increased** thanks to him,

7B	*foruz-ān*	***shod-e***	*takht=e*	*shāhi*	*bed-u*
	shine.PRS-PTCP	become.PST-PTCP	throne=EZF	royal	thanks.to-him

the royal throne [having] **become** resplendent because of him.

[18] Ferdowsi, *The Shahnameh*, vol. 1, p. 41 (*Jamshid*, lines 4-7).

The translation tries to render the structure of the four lines in question, which represent a single sentence. Each half-line is a syntactic unit with a verb in the past participle or without an explicit verb.

The syntactic connection between the finite clause (He [Jamshid] ascended the throne) in the first half-line, and the following syntactic units, as well as the connection of the latter units with each other, is very loose. All syntactic units have a subject differing from the subject of the finite clause, with the exception of line 5A, in which the subject of *baste* is 'Jamshid', as in the finite clause.[19] The two nominal sentences 4B (*bar sar-ash tāj-e zar* 'on his head [there being] a golden crown'), and 6B (*be farmān-e u div o morgh o pari* '[being] at his orders the demons, the winged beings and the fairies') are coordinated with the participial constructions and some form of 'to be' is probably implied.

One could be led to interpret the passage in example [5] as composed of a series of sentences asyndetically coordinated to the clause with a finite verb placed at the beginning. However, the form of the past participle in the syntactic units that follow the clause with a finite verb seems to prove that a different syntactic function is involved.[20] The formal difference between the past finite form (*bar āmad* 'he went up') in the clause placed at the beginning, and the participial forms (*baste*, *gashte*, etc.) in the following syntactic units, suggests that the syntactic units with a participle are not coordinated with the finite clause.[21] However, though the participles signal a relation of dependency, the syntactic units that follow the finite clause cannot be really considered as subordinated either: they are asyndetically juxtaposed, without coordinating or subordinating conjunctions, to the finite clause, with alternation of same-subject and different-subject participles. In Persian literature, the construction described above is mainly used for stylistic

[19] *Kamar baste* has been considered here as a past participle from the periphrastic verb *kamar bastan* 'to get ready for action', lit. 'fasten the belt', rather than as a lexicalized compound adjective: *kamar-baste* 'prepared, ready for action; obedient to orders' (for which see Anvari, *Farhang*, vol. 6, p. 5929). Both readings, however, are possible.

[20] The meter in line 7 confirms the reading *fozude* and *shode* (past participles), instead of *fozud* and *shod* (preterit, 3rd sg), as the ending *-e* of the past participle in the above forms counts as a long syllable. Consequently, the other coordinated verbal forms, too, must be past participles. Cf. the remarks by Khaleghi Motlagh, 2001-2009, vol. 1, part 1, p. 49.

[21] Cf. the notion of 'deranking' as formulated by Shagal (2019: 38-41) about non-finite, or deranked, forms such as participles: "[I]n deranked constructions the predicate of the subordinate clause exhibits structural differences from the main clause predicate" (Shagal 2019: 38). Accordingly, the author defines participles as "deranked verb forms that can be employed for adnominal modification" (Shagal 2019: 1, 52).

purposes in descriptive passages in narrative poetry. In prose, constructions with some similarities with the constructions attested in the *Shāhnāme* are not infrequent (see in particular example [15] below, from a prose historical text), which shows that such participial constructions may have belonged to the ordinary language as well.

Cases in which the participial construction precedes the finite verb are occasionally found in the *Shāhnāme*:

[6]

همه دست **بر داشته** باسمان همی خواندندش به نیکی گمان.[22]

hame	*dast*	***bar***	***dāsht-e***	*b-āsmān*
everyone	hand[OBJ]	up	keep.pst-ptcp	to-sky

Everybody, raising his hands to the sky,

hami	*khʷānd-and=ash*	*be*	*niki-gomān*
dur	acclaim.PST-3PL=him	as	well-wisher

was acclaiming him as a well-wisher.

Here the participial construction *dast bar dāshte b-āsmān* (=*be āsmān*) 'raising/ having raised [their] hands to the sky', co-referential with the subject of the main clause (*hame* 'everyone'), is embedded between the subject and the finite verb form, which is in the past continuous tense (*hami* + past tense). In this case, the participial construction shows a clearer subordinate syntactic behaviour (on extraposition and embedding as characteristic of subordinate clauses see Weisser 2015: 11-14. See also example [14] below).

3. Examples from earlier poetry texts

According to Djalal Khaleghi Motlagh, constructions with a past participle (*vajh-e vaṣfi*) – particularly frequent in the section of the *Shāhnāme* that Ferdowsi wrote in his youth – are already frequent in the nearly thousand lines by Daqiqi (second half of the tenth century) incorporated into the text of Ferdowsi's poem.[23] In fact, instances of such past participles from Daqiqi's

[22] Ferdowsi, *The Shahnameh*, vol. 1, p. 91 (*Feridun* line 38).

[23] Cf. Djalal Khaleghi Motlagh, 2001-2009, vol. 1, part 1, p. 49. In Khaleghi Motlagh's edition, the incorporated lines by Daqiqi are the following: Ferdowsi, *The Shahnameh*, vol. 5, pp. 76-174 (*Goshtāsp*, lines 14-1028). A thorough study of

verses are numerous. For example, when King Goshtāsp dismisses the ambassadors sent by Arjāsp and sends back with them a threatening letter that Zarir, Goshtāsp's brother, had written for Arjāsp, Daqiqi recounts:

[7]

231 فرستادگان سپهدار چین زِ پیشِ جهاندار شاهِ زمین،
232 برفتند هر دو **شده** خاکسار جهاندارشان **رانده** و **کرده** خوار [...]
234 چو از دور دیدند ایوانِ شاه **زده** بر سر او درفش سیاه،
235 فرود آمدند از چمنده ستور شکسته‌دل و چشم‌ها **گشته** کور
236 پیاده برفتند تا پیشِ اوی سیه پاکشان جامه و زرد روی
237 بدادندش آن نامه ی شهریار به پاسخ **نبشته** زریرِ سُوار.[24]

231A *ferestādeg-ān=e* *sepahdār=e* *Chin*
envoy-PL=of general=of China

The envoys of the general of China [Arjāsp],

231B *ze* *pish=e* *jahān-dār* *shāh=e* *zamin*
from in.front=of world-keeper king=of earth

from the presence of the king keeper of the world [Goshtāsp]

232A *be-raft-and* *har* *do* ***shod-e***
PFV-go.PST-3PL every two become.PST-PTCP

khāksār
covered.with.dust

departed, both [being/ having been] **humiliated**,

232B *jahān-dār=eshān* ***rānd-e*** *vo*
world-keeper[SBJ]=them[OBJ] expel.PST-PTCP and

kard-e *khār […]*
do.PST-PTCP base […]

the king [having] **cast** them **out** and **abased** them[…].

234A *cho* *az* *dur* *did-and* *eyvān=e* *shāh*
when from afar see.PST-3PL palace[OBJ]=of king

When they saw the king's [i.e. Arjāsp's] palace in the distance,

Daqiqi's incorporated lines, from a stylistic and lexical perspective, is offered by Khāleqi-Motlaq 2002. From the historical-literary point of view see Dahlén 2011.

[24] Ferdowsi, *The Shahnameh*, vol. 5, p. 98 (*Goshtāsp*, lines 231-232, 234-237).

234B	***zad-e***	*bar*	*sar=e*	*u*	*derafsh=e*	*siyāh*
	place.PST-PTCP	on	top=of	it	banner=EZF	black

– a black banner [having been] **placed** on its top –

235A	*forud*	*āmad-and*	*az*	*cham-ande*
	down	come.PST-3PL	from	walk.gracefully.PRS-PTCP

sotur
horse

they dismounted from the steeds,

235B	*shekaste-del*	*o*	*chashm-hā*	***gasht-e***
	broken-hearted	and	eye-PL[SBJ]	become.PST-PTCP

kur
blind

[with] an afflicted heart, their eyes **turned** blind.

236A	*piyāde*	*be-raft-and*	*pish=e*	*uy*
	on.foot	PFV-go.PST-3PL	in.front=of	him

On foot they presented themselves before the king,

236B	*siyah*	*pāk=eshān*	*jāme*	*o*	*zard*	*ruy*
	black	pure=POSS.3PL	robe[SBJ]	and	yellow	face[SBJ]

their pure robes [turned] black, and their faces [turned] yellow.

237A	*be-dād-and=sh*	*ān*	*nāme=ye*	*shahriyār*
	PFV-give.PST-3PL=them	that	letter[OBJ]=of	king

They handed him the king's [i.e. Goshtāsp's] letter,

237B	*be*	*pāsokh*	***nebesht-e***	*Zarir=e*	*sovār*
	in	response	write.PST-PTCP	Zarir=the	knight

[which] Zarir the knight [had] **written** in response.

In this passage from Daqiqi's text, each of the four finite verbs is followed by one or more syntactic units with a verb in the past participle (see below, Units 1-4). The subject of the finite verbs is the same in all cases, i.e. the two envoys sent by Arjāsp. In one case, the clause with a finite verb is preceded by a time clause: *cho az dur didand eyvān-e shāh* 'When they saw the king's [Arjāsp's] palace in the distance' (234A). The latter is followed by a 'parenthetical' absolute participial construction (234B), referring to the king's palace: 'on its top a black banner [had been/was] placed (*zade*)'.[25] After which

[25] *Zade* '[having been] placed' is used in a passive meaning, without an agent expressed. This must be linked to the semantics of the verb *zadan* 'to strike'. In

the narration is resumed: *forud āmadand az chamande sotur* 'they dismounted from the steeds' (235A):

Unit 1

1.	*beraftand*	[The two envoys] departed
1.a.	*har do* ***shode*** *khāksār*	both [having been] humiliated
1.b.	*jahāndār-eshān* ***rānde*** *vo* ***karde*** *khār*	the king [having] driven them out and abased them.

Unit 2

2.	*forud āmadand*	They dismounted [from the steeds]
2.a.	*shekaste-del*[26] *o chashmhā* ***gashte*** *kur*	with an afflicted heart, their eyes [having] turned blind.

Unit 3

3.	*piyāde beraftand pish-e uy*	On foot they went before him,
3.a.	*siyah pāk-eshān jāme o zard ruy*	their pure robes [turned] black, and their faces yellow.

Unit 4

4.	*bedādand-sh ān nāme-ye shahriyār*	They handed him the king's [i.e. Goshtāsp's] letter
4a.	*be pāsokh* ***nebeshte*** *Zarir-e sovār*	[which] Zarir the knight [had] written in response.

Absolute participles and participles having the same subject as that of the governing verb alternate in this as in the above examples. Indeed, the participles in Units 1.b and 2.a have a subject differing from that of the

compound adjectives formed with a past participle, *zade* often has a passive meaning: *shegeft-zade* 'stricken by astonishment', *āfat-zade* 'stricken by misfortune', etc.

[26] *Shekaste-del* 'broken-hearted' is a lexicalized compound adjective in which the past participle *shekaste* 'broken', from *shekastan* 'to break', has a passive or resultative value: '[someone] whose heart has been/ is broken'. See Anvari, *Farhang,* vol. 5, p. 4539.

governing finite verb, as the referent of the subject no longer is the two envoys, but in 1.b it is the king of the world (i.e. Goshtāsp), and in 2.a the envoys' eyes. In Unit 4, the clause with a finite verb, *bedādand-sh (=ash) ān nāme-ye shahriyār* 'They handed him the king's letter', is followed by a relative participial construction, '[which] Zarir the knight [had] written (*nebeshte*) in response', that modifies the object (the king's letter) of the finite clause (line 237). However, the past participle *nebeshte* 'written' from *nebeshtan* 'to write', can also be interpreted as endowed with a passive value. In this case, 'Zarir' would be the unmarked agent: '[the letter] written in response [**by**] Zarir the knight'.

The usage of participial constructions, already attested in the verses by Daqiqi incorporated into the *Shāhnāme*, is therefore ancient. In an attempt to go back to the very first attestations of this syntactic construction in Persian literary poetry, I examined the first 50 pages of Gilbert Lazard's edition of the most ancient fragments of Persian poetry. In these nearly 300 lines, despite the fragmentary form in which they have reached us, some constructions with a past participle can be identified.[27] Among them there is a line by Firuz Mashreqi (d. 283/896, a contemporary of the Saffarid ʿAmr-e Leyth):

[8]

نوحه گر کرده زبان چنگ حزین از غم گل موی **بگشاده** و بر روی زنان ناخونا.[28]

nowhegar	*kard-e*	*zabān*	*chang=e*	*ḥazin*	*az*
mourner	do.PST-PTCP	tongue[OBJ]	harp[SBJ]=EZF	afflicted	from

gham=e	*gol*
pain=of	rose

The harp, suffering from pain for the rose, has made its tongue a professional mourner,

[27] Sure examples of participial constructions can be found in Lazard 1964: vol. 2, p. 19 (Firuz Mashreqi, lines 2, 4); vol. 2, pp. 29-31, 34 (Shahid Balkhi, lines 41, 46, 55, 79).

[28] Lazard 1964 vol. 2, p. 19. Lazard more freely translates: "La harpe attristée par la rose élève une voix gémissante; – la chevelure éparse, elle lacère de ses ongles son visage" (1964: vol. 1, p. 60). For images referring to the complaint of musical instruments, see Beelaert 2000: 181-198.

muy	***be-gshād-e***	*vo*	*bar*	*ruy*	*zan-ān*
hair[OBJ]	PFV-untie.PST-PTCP	and	on	face	hit.PRS-PTCP

nākhun=ā
fingernail[OBJ]=INTERJECTION

[having] **untied** its hair, and scratching its face with its fingernails.

In the absence of a context, it is possible that in this line, which is the first of a two-line fragment, the past participle *karde* is a perfect with an implied *ast* (see above, § 1), and has been translated as such. In contrast, *muy* ***begshāde*** (=*begoshāde)* '[having] untied its hair' is to be considered a participial construction endowed with adverbial value, expressing the manner of the action: the harp was complaining '[having] untied [its] hair/ [with] untied hair'. The coordinated (*vo=va* 'and') clause which follows, *bar ruy* ***zanān*** *nākhun* 'hitting with its nails on its face', with a present participle, also functions adverbially.

4. Examples from early prose texts

Constructions with a past participle in the same functions as the ones described above are to be found in early prose texts too, as the following examples show:

[9]

و مردی از در اندر آمد چوبی بدست **گرفته**.[29]

va	*mard=i*	*az*	*dar*	*andar*	*āmad*	*chub=i*
and	man=ART.INDF	from	door	in	come.PST[3SG]	club=ART.INDF

be	*dast*	***gereft-e***
in	hand	take.PST-PTCP

A man came in [after having] **taken** a club in his hand/ A man [who had] **taken** a club in his hand came in.

In this example, from the History of Balʿami (second half of the tenth century), the past participle *gerefte* '[having] taken', with a converb-like, or possibly with a relativizing function, is referentially controlled by the subject

[29] Balʿami, *Tārikh*, ed. M.-T. Bahār and M. Parvin-Gonābādi, vol. 2, p. 1094.

of the superordinate clause (a man). It has an active transitive value and governs an indefinite non-marked object (a club). It follows the clause with a finite verb ('a man came in'), like in the poetry examples discussed above.

[10]

و[ابو علی سینا] با تنی دو در رفت و کاردی به دست **گرفنه** گفت...[30]

va	*[Abu ʿAli Sinā]*	*bā*	*tan=i*	*do*	*dar*	*raft*
and	[Avicenna]	with	people=ART.INDF	two	inside	go.PST[3SG]

va	*kārd=i*	*be*	*dast*	***gereft-e***	*goft…*
and	knife=ART.INDF	in	hand	take.PST-PTCP	say.PST[3SG]

[Avicenna] with a couple of people came in and, [after having] **taken** a knife in his hand, said…

In this example, from a later text (mid-twelfth century), the same participial construction ('having taken [something] in [his/her] hand') precedes the finite verb ([Avicenna] said) and has a more marked subordinate adverbial function.

[11]

بهرام بفرمود تا ... همچنان کمان بزه **کشیده** بر پشت اسپ و آن گور و شیر و تیر اندر زمین همچنان صورت کردند.[31]

Bahrām	*be-farmud*	*tā […]*	*hamchonān*	*kamān*	*be*	*zeh*
Bahrām	PFV-order.PST[3SG]	that […]	still	bow	to	string

kashid-e	*bar*	*posht=e*	*asp*	*va*	*ān*	*gur*	*o*	*shir*
drow.PST-PTCP	on	back=of	horse	and	that	onager	and	lion

o	*tir*	*andar*	*zamin*	*hamchonān*	*ṣurat*	*kard-and*
and	arrow	into	ground	just.that.way	portrait	do.PST-3PL(generic)

Bahrām ordered that […] they make his portrait just that way, while he was still drawing his bow on horseback, and that onager, the lion, and the arrow [with which he had hit them in a single shot] on the ground.

In this example, taken, like example [9] above, from the History of Balʿami, *kashide* 'drawn' can be interpreted as an active transitive past participle governing a non-marked object (the bow), employed in a converb-like function. It is co-referential with the implicit object (King Bahrām) in the

[30] Neẓāmi ʿArużi, *Chahār maqāle*, ed. M. Qazvini, p. 83. Edward G. Browne's translation of the passage is: "Taking a knife in his hand, he entered with two attendants, saying…" (*Four Discourses*, p. 127).

[31] Balʿami, *Tārikh*, ed. M.-T. Bahār and M. Parvin-Gonābādi, vol. 2, pp. 930-931.

superordinate purpose clause: Bahrām ordered [them] to depict [**him**] while he was drawing his bow on horseback, exactly the way his famous hunting deed was accomplished. However, the possibility of interpreting *kashide* as endowed with a passive meaning, and the participial construction *kamān be zeh kashide* as an absolute construction is also possible. In this case *kamān* 'bow' would be the subject of the participle (the bow having been drawn).

[12]

دزدان از سه جانب بر خاستند و سوی کاروان در آمدند شمشیرها **کشیده**.[32]

dozd-ān	*az*	*se*	*jāneb*	*bar*	*khāst-and*	*va*	*su=ye*
robber-PL	from	three	side	up	rise.PST-3PL	and	direction=of

kārvān	*dar*	*āmad-and*	*shamshir-hā*	***kashid-e***
caravan	out	come.PST-3PL	sword-PL	drow.PST-PTCP

The robbers rose up on three sides and ran towards the caravan with drawn swords.[33]

As in the just discussed example [11], in this example too, from the *Siyar al-Moluk* (end of the eleventh century), the past participle *kashide* '[having] drawn' has the function of a converb. It has been interpreted as an active transitive participle referentially controlled by the subject of the superordinate clause (the robbers), and governing a generic non-marked object (the swords): the robbers had drawn their swords. Otherwise, it can be interpreted as part of an absolute construction, whose subject would be *shamshirhā* 'the swords': the robbers rose up, the swords [having been, being] drawn (see example [11] above).[34] As in the *Shāhnāme* examples, the participial construction follows the finite clause.

[13]

چون بکاروان رسیدند همهٔ صحرا مردم **مرده** دیدند و سپر و شمشیر و زوبین و تیر کمان **بیفگنده**.[35]

[32] Neẓām al-Molk, *Siyar al-moluk*, ed. H. Darke, p. 93.

[33] Neẓām al-Molk, *The Book of Government,* transl. by H. Darke, p. 69.

[34] The past participle *kashide* is lexicalized as an adjective with a passive meaning: 'unsheathed'. See Anvari, *Farhang,* vol. 6, p. 5841: "nr. 8 (adj., archaic) 'taken out of the sheath'".

[35] Neẓām al-Molk, *Siyar al-moluk*, ed. H. Darke, p. 94.

chun	*be*	*kārvān*	*rasid-and*	*hame=ye*	*ṣaḥrā*	*mardom*
when	to	caravan	arrive.PST-3PL	all=of	plain	people[OBJ]

mord-e	*did-and*	*va*	*separ*	*o*	*shamshir*	*o*	*zubin*
die.PST-PTCP	see.PST-3PL	and	sheld	and	sword	and	javelin

o	*tir*	*<o>*	*kamān*	***biy-afgand-e***
and	arrow	<and>	bow	PFV-strew.PST-PTCP

When [the emir and his men] reached the caravan, [across] the entire plain they saw people **dead**, and shields, swords, javelins, bows and arrows **strewn** around.

In this example, likewise from the *Siyar al-moluk*, the participle *morde* 'dead' is referentially controlled by the object ('people', i.e. the robbers) of the verb *didand* 'they saw'. It is employed in a predicative function: the emir and his men saw the robbers **dead**. The second participle, *biyafgande* 'strewn', from the transitive verb *afgandan* 'to throw, strew', is likewise interpretable as a copredicative adjective, endowed with a passive-resultative meaning ('fallen, strewn'): they saw the shields and weapons **strewn** around.[36] But the past participle *biyafgande* can be also interpreted as endowed with an active verbal function: the emir and his men saw the dead people who [before dying] had abandoned their shields and weapons across the plain.[37]

[14]

ای فلان چرا نان نیم **خورده** از خوان ما بر خواستی.[38]

ey	*felān*	*cherā*	*nān*	*nim*	***khword-e***	*az*
INTERJECTION	So-and-so	why	meal	half	eat.PST-PTCP	from

khwān=e	*mā*	*bar*	*khāst-i*
table=of	us	up	rise.PST-2SG

Hey you, why, with the meal only half **eaten**, did you leave our table?[39]

---

[36] Herbert Darke's translation of the passage corresponds to this interpretation: "When they reached the caravan they saw the plain strewn with corpses, shields, swords, clubs, bows and arrows" (Neẓām al-Molk, *The Book of Government*, p. 70). For the past participle *afkande* in the intransitive meaning of 'fallen, strewn', see Anvari, *Farhang,* vol. 1, p. 499.

[37] The latter is the interpretation reflected in the Italian translation by M. Pistoso: "… videro l'intera pianura disseminata di morti **che** avevano abbandonati scudi, spade, archi e frecce e giavellotti" (Neẓām al-Molk, *L'arte della politica*, p. 132).

[38] *Qābus-nāme*, ed. Gh.-Ḥ. Yusofi, p. 65.

[39] Translation adapted from Key Kāʾus b. Eskandar, *A mirror for princes*, transl. by Reuben Levy, p. 56.

The past participle *khworde* 'eaten' in the participial construction *nān nim khworde* '[having] eaten half [your] meal' functions as a converb, co-referential with the subject of the superordinate clause (you). It has an active transitive meaning and governs a generic, non marked object (*nān*, 'bread, meal'). In this case, too, it is possible to interpret instead the participial construction as an absolute construction having *nān* 'bread, meal' as its subject, differing from the subject (you) of the superordinate clause (the meal [having been] half eaten). In any case, in this example the participial construction – unlike the examples from the *Shāhnāme* and other early poetry and prose texts – is embedded into the main clause, and has a clear subordinate value (see also example [6] above).

[15]

و ما درین هفته از اینجا حرکت خواهیم کرد همه مرادها حاصل **گشته** و جهانی در هوا و طاعت ما **بیارامیده**.[40]

va	*mā*	*dar*	*in*	*hafte*	*az*	*injā*	*ḥarakat*	*khwāh-im*
and	we	in	this	week	from	here	move	want.PRS-1PL

kard	*hame*	*morād-hā*	*ḥāṣel*	***gasht-e***	*va*
do.INF	all	intention-PL[SBJ]	achieved	become.PST-PTCP	and

jahān=i	*dar*	*havā*	*va*	*ṭāʿat=e*	*mā*
world[SBJ]=ART.INDF	in	love	and	obedience=of	us

biy-ārāmid-e
PFV-rest.PST-PTCP

During this week we intend to leave from here, all our intentions [having been] **achieved** and an [entire] world [being] **pacified** under our desire and command.[41]

Unlike examples [11-14] above, the two participial constructions in this passage, with past participles from intransitive verbs, can only be interpreted as absolute constructions, each having its own subject (*morādhā* 'intentions', *jahān-i* 'an [entire] world'), differing from the subject of the superordinate clause (we). The two absolute constructions follow the clause with a finite verb and recall the constructions in example [5] above.

[40] Abo'l- Fażl Beyhaqi, *Tārikh-e Beyhaqi*, ed. Ghani and Fayyāż, p. 83

[41] In Abo'l- Fażl Beyhaqi, *The History of Beyhaqi* (transl. by C.E. Bosworth and M. Ashtiany, vol. 1, p. 165) the translation is as follows: "We ourselves in the course of this week intend to make a move from here, having achieved all our desires and with the people secure and content under our rule and command".

Predicative constructions, with (or without) a past participle, governed by verbs of sense (e.g. *didan* 'to see', or *yāftan* 'to find, see as'; see examples [3-4] and [13]) are destined to great fortune in New Persian poetry and prose especially in descriptions, as the following example from the *Tadhkerat al-ʿOwliyā* by ʿAṭṭār (end of the eleventh–beginning of the twelfth century) shows:

[16]

مردی دیدم زرد روی نحیف **شده** چشم در مغاک **افتاده**.[42]

mard=i	*did-am*	*zard-ruy*	*naḥif*	***shod-e***
man[OBJ]=ART.INDF	see.PST-1SG	yellow-faced	weak	become.PST-PTCP

chashm	*dar*	*maghāk*	***oftād-e***
eye[SBJ]	in	socket	sink.PST-PTCP

I saw a man yellow on cheeks, emaciated, his eyes **sunken** into his sockets.

In Neẓāmi's poem *Khosrow and Shirin*, from the second half of the twelfth century, the description of Shirin bathing in the spring, as seen by Khusraw's passionate eyes, can offer a further example of the use of past participles in descriptions. In this example, different semantic and functional nuances of past participles are represented, from the Arabic participle *mohayyā* 'ready, prepared' and of *neshaste* 'sitting', both employed as adjectives in a predicative function, to the adverbial or relativizing function of *baste* '[after] having tied/ [who had] tied':

[17]

عروسی دید چون ماهی مهیا […]

در آب نیلگون چون گل **نشسته** پرندی نیلگون تا ناف **بسته**.[43]

ʿarus=i	*did*	*chun*	*māh=i*	*mohayyā* […]
bride[OBJ]=ART.INDF	see.PST[3SG]	like	moon=ART.INDF	prepared […]

He saw a bride prepared as a moon […]

dar	*āb=e*	*nilgun*	*chun*	*gol*	***neshast-e***
in	water=EZF	blue	like	flower	sit.PST-PTCP

lying on the blue water like a flower,

[42] Farid al-Din ʿAṭṭār, *The Tadhkiratu ʾl-Awliyá*, ed. Reynold A. Nicholson, Part I, p. 117.

[43] Neẓāmi, *Xosrow va Shirin,* ed. B. Tharvatiyān, pp. 190-191, ch. 24, lines 42A and 44.

parand=i	*nilgun*	*tā*	*nāf*	***bast-e***
cloth[OBJ]=ART.INDF	blue	up.to	navel	tie.PST-PTCP

[having/ who had] **tied** a blue cloth up to the navel.

5. The negative form

In the negative form the participle is generally preceded, at least in early texts, by the negative particle *nā-*. An example from the *Shāhnāme* is found in an episode narrating an embassy from the Qeyṣar, the Byzantine emperor, to Khosrow Anushervān. The Qeyṣar challenges the sages of Khosrow Anushervān's court, saying:

[18]

که با شاهِ گنداوران و ردان فراوان بود پاک دل موبدان

بدین درج و این قفل **نابرده دست** نهفته بگویند چیری که هست.[44]

ke	*bā*	*shāh=e*	*gondāvar-ān*	*o*	*rad-ān*
that	to	king=of	brave.man-PL	and	learned-PL

"The King of the brave and the wise

farāvān	*bov-ad*	*pākdel*	*mowbad-ān*
many	be.PRS-3SG	pure.hearted	priest-PL

has many priests with a pure heart.

bed-in	*dorj*	*o*	*in*	*qofl*	***nā-bord-e***	*dast*
in-this	casket	and	this	lock	NEG-take.PST-PTCP	hand

Without touching this locked casket,

nehoft-e	*be-guy-and*	*chiz=i*	*ke*	*hast*
hide.PST-PTCP	SBJV-say-3pl	thing=DET	REL	be.PRS[3SG]

let them say what is hidden in there".

In the following example, after Fereydun defeats and tightly ties up Żaḥḥāk, it is recounted:

[44] Ferdowsi, *The Shahnameh*, vol. 7, p. 371 (*Nushin Ravān*, lines 3603-3604).

[19]

دُمادُم برون رفت لشکر ز شهر وُزان شاه **نایافته** شهر بهر.[45]

dom-ā-dom	*berun*	*raft*	*lashkar*	*ze*	*shahr*
one.after.another	out	go.PST[3SG]	army[SBJ]	from	city

In an orderly line the army left the city,

vo-z	*ān*	*shāh*	***nā-yāft-e***	*shahr*	*bahr*
and-from	that	king	NEG-find.PST-PTCP	city[SBJ]	advantage[OBJ]

without the city having been able to take advantage of that king.

Other examples of participles in the negative form are the following, from early prose texts:

[20]

و هر که بدین مقام **نارسیده** قدم آنجا نهد زندیق و اباحتی و کشتنی بود.[46]

va	*har-ke*	*bed-in*	*maqām*	***nā-rasid-e***	*qadam*
and	everyone-who	to-this	state	NEG-arrive.PST-PTCP	footstep

ānjā	*neh-ad*	*zendiq*	*va*	*ebāḥati*	*va*
there	put.PRS-3SG	disbeliever	and	ungodly	and

koshtani	*bov-ad*
destined.for.slaughter	be.PRS-3SG

Anyone who ventures there without having attained that level is a disbeliever and ungodly and deserves death.

[45] Ferdowsi, *The Shahnameh*, vol. 1, p. 84 (*Żaḥḥāk*, line 471). At the beginning of the second half-line (471B) the conjunction *va* maintains its full value, with a concessive nuance (on the so-called *vāv-e ḥāliye* see Shafiʿi 1964: 353; Farshidvard 1999: 323-324): 'and [=though] the city having not been able to take advantage of that king', i.e. the people of the city had had only a little time to watch Fereydun defeating and tightly tying Żaḥḥāk (for this interpretation cf. Khaleghi Motlagh 2001-2009, vol. 1, part 1, p. 105). For line 471B different readings are given by manuscripts. The same line is quoted by Farshidvard (1999: 325) according to what seems to be a *lectio facilior*: *vo-z-ān shahr nā-yāfte hich bahr* 'without [the army] having taken any advantage of [the pillaging of] the city'.

[46] Farid al-Din ʿAṭṭār, *The Tadhkiratu ʾl-Awliyá*, ed. Reynold A. Nicholson, Part I, 122-123.

[21]

هرگز کسی به جهل خویش اقرار نکند مگر آن کس که چون دیگری در سخن باشد [همچنان] تمام **ناگفته**، سخن آغاز کند.[47]

hargez	*kas=i*	*be*	*jahl=e khwish*	*eqrār*
ever	person=ART.INDF	to	ignorance=POSS.REFL	attestation

na-kon-ad	*magar*	*ān*	*kas*	*ke*	*chun*	*digar=i*
NEG-do.PRS-3SG	except	that	person	REL	when	other=ART.INDF

dar	*sokhan*	*bāsh-ad*	*[hamchonān]*	*tamām*	***nā-goft-e***
in	discourse	be.PRS-3SG	[yet]	entirely	NEG-talk.PST-PTCP

sokhan	*āghāz*	*kon-ad*
discourse	beginning	do.PRS-3SG

No one ever reveals his ignorance except the one who, when another is talking, starts talking without waiting for him to finish.

As *nā-* is a prefix used for the negative form of nouns and adjectives (participles included), its usage before past participles in the examples above is a further proof that in the syntactic constructions in question the past participle does not represent a perfect form with an implied auxiliary, because – in such case – the negative prefix would have been *na-* (Lazard 1963: 442-3 §§ 730-731).

6. Conclusions

The present study describes participial constructions from an early narrative poem, the *Shāhnāme*, compared with more or less coeval prose texts. They are generally placed after a clause with a verb of finite form in the past and have a general 'descriptive' value, in as much as they, roughly speaking, correspond to a qualifying, an adverbial or a predicative expression. Indeed, in the analyzed examples, past participles are used in three strictly related and often indistinguishable syntactic and semantic functions:

1. as an adnominal (relatiziving) modifier, or as the verb of a relative clause lacking a relative pronoun and an auxiliary;
2. as an adverbial modifier, or as the verb of an adverbial clause;
3. as an adjective – or as part of an adjective-, noun-, or prepositional phrase – employed in a predicative function.

[47] Saʿdi, *Golestān*, ed. Gh.-Ḥ. Yusofi, 130 (Ch. 4, *hekāyat* 7).

The position of participial constructions after the finite verb may correspond to the word order generally attested for Early and Classical New Persian. Indeed, in Early and Classical New Persian, nominal expressions employed in a qualifying, predicative, or adverbial function were preferably placed after the verb, as in *maqām-i did del-goshā* 'he saw a pleasant place', lit. 'a place he saw [that was] pleasant' (see Bausani 2023: 56-59, 70-73 [12-15, 26-29], with other examples).[48] The comparison with the word order characteristic of (literary) Early New Persian prose could reinforce the interpretation of participial constructions as (mainly) nominal syntactic units. However, any attempt to separate too rigidly the nominal from the verbal value of past participles in early constructions is doomed to failure. This said, from the analysis above a few remarks can be made.

Functions 1 and 2 are mainly distinguishable on the basis of the semantic context. In these functions, past participles indicate a time relationship of anteriority in comparison with the action expressed by the finite verb. Past participles from transitive verbs generally have an active meaning and may govern either a generic or a specific object. They can be co-referential with either the subject or the object of the finite clause verb. If the subject of the past participle is the same as that of the finite clause (same-subject reference), the subject of the participle is normally left implicit. Otherwise, the subject of the past participle is a (necessarily expressed) different subject. The latter case represents what is called an 'absolute construction'. Given the fact that Persian does not possess a nominal inflection (apart from plural marking), and that subject and unmarked object are morphologically indistinguishable, both interpretations are possible in the case of past participles from transitive verbs also endowed with a lexicalized intransitive meaning (*kashide*, *afgande*, etc.). In these cases, the past participle can have an active transitive meaning, governing an unmarked object, or an intransitive resultative meaning, with the noun or noun phrase functioning as the subject of the absolute construction (see examples above, in particular [11-14]).

In functions 2 and 3 past participles syntactically depend on the verb of the main clause. As for their reference, they can be co-referential with either the subject or the object of the main clause. In function 3, they describe a state or a quality ascribed to the subject or the object of the finite verb, and have the value of a predicative expression. Just like participles in function 2, they often function much like converbs. In this regard, note that

[48] Bausani (2023: 72 [=28]) considers such post-verbal determinants as the issue of Middle Persian relative clauses with dropping of the relative pronoun/ezafe particle *ī*, as in *pus-ē ast ī pad frahang ud aswārīh frahixtag* 'he is a boy who [is] educated in knowledge and in riding', in an example from the *Kārnāmag ī Ardašīr ī Pābakān*. On determinative participial constructons in Middle Persian see Asatrian 1989: 28.

in predicative expressions without a past participle, some form of 'to be' or other linking verbs employed in a converb-like function can be implied.

In the *Shāhnāme*, absolute participial constructions, also known from early prose texts, have a characteristic development in narratives, as a means to produce a lively and stylistically swift description. Often they form long clause chains conveying a sequence of events, in which the past participle signals a relation of dependency from the clause of finite form.

In the history of the Persian language only the adverbial function of participial constructions survives up to the present. In the other functions, either relative clauses are used, such as **mard-i-rā didam* ***ke zard-ruy va naḥif bud*** 'I saw a man who was yellow on cheeks and emaciated' (cf. example [16] above), or nominal groups without a participle (very often, noun + prepositional phrase), as in ***dast be sine*** *kenār istād* 'a hand on [his] chest he stood aside' (for these expressions see Lazard 2006: 190 § 189). The development of a clearly subordinated participial construction with an exclusively adverbial value may have occurred as a consequence of the new syntactic position of participial constructions, which over time became fixed before the finite clause. It is the position normally occupied in Persian by adverbial clauses of time, cause, manner, condition.

BIBLIOGRAPHY

PRIMARY SOURCES

Anvari, Ḥasan et alii. *Farhang-e Sokhan*. Tehrān, 1381/2002, 8 vols.

ʿAṭṭār, Farid al-Din. *The Tadhkiratu ʾl-Awliyá*. Part I. Ed. by Reynold A. Nicholson. London: Luzac & Leiden: Brill, 1905.

Balʿami, Abu ʿAli Moḥammad b. Moḥammad b. Balʿami. *Tārikh-e Balʿami. Takmele va tarjome-ye Tārikh-e Ṭabari*. Ed. by Moḥammad-Taqi Bahār "Malek al-Shoʿarā". Revised edition by Moḥammad Parvin-Gonābādi. 2nd ed. 2 vols. Tehrān: Zavvār, 1353/1974.

Beyhaqi, Aboʾl- Fażl, *Tārikh-e Beyhaqi*. Ed. by [Qāsem] Ghani and [ʿAli-Akbar] Fayyāż. Tehrān: Chāpkhāne-ye Bānk-e melli-ye Irān, 1324/1945.

Beyhaqi, Aboʾl- Fażl: *The History of Beyhaqi (the History of Sultan Masʿud of Ghazna, 1030-1040) by Abuʾl-Fażl Beyhaqi*. Translated with a historical, geographical, linguistic and cultural commentary and notes by C.E. Bosworth. Fully revised and with further commentary by Mohsen Ashtiany. 3 vols. Boston & Washington: The Ilex Foundation & Center for Hellenic Studies Trustees for Harvard University, 2011.

Ferdowsi, Abuʾl-Qasem. *The Shahnameh (The Book of Kings)*. Ed. by Dj. Khaleghi-Motlagh with M. Omidsālār and A. Khaṭibi. 8 vols. Albany: University of New York, 1987-2008.

Key Kāʾus b. Eskandar, *Qābus-nāme*. Ed. by Gholām-Ḥoseyn Yusofi. Tehrān: Bongāh-e tarjome va nashr-e ketāb, 1345/1966.

Key Kāʾus b. Eskandar, *A mirror for princes. The Qābūs-nāma by Kai Kāʾūs ibn Iskandar Prince of Gurgan*. Transl. by Reuben Levy. New York: E.P. Dutton & Co., 1951.

Neẓām al-Molk, *Siyar al-moluk*. Ed. by H. Darke, 2nd ed. Tehrān: Bongāh-e tarjome va nashr-e ketāb, 1347/1962.

Neẓām al-Molk, *The Book of Government or Rules for Kings. The Siyar al-Muluk or Siyasat-Nama of Nizam al Mulk*. Transl. from the Persian by Hubert Darke. New York: Rouledge, 2002 [1st ed. 1960].

Neẓām al-Molk, *L'arte della politica (Lo specchio del principe nella Persia dell'XI secolo). …* Traduzione dal persiano con introduzione e note a cura di Maurizio Pistoso. Milano & Trento: Luni Editrice, 1999.

Neẓāmi ʿArużi, *Chahār maqāle*. Ed. by Moḥammad Qazvini. Leiden: Brill, 1909.

Neẓāmi ʿArużi, *Chahār maqāle*. Revised translation of the Chahár Maqála ("Four Discourses") […] by Edward G. Browne. Cambridge & London: Cambridge University Press & Luzac, 1921.

Neẓāmi Ganjayi. *Xosrow va Širin*. Ed. by Behruz Tharvatiyān. Tehrān: Ṭus, 1366/1987.

Saʿdi, *Golestān*. Ed. by Gholām Ḥoseyn Yusefi, 2nd ed. Tehrān: Khwārazmi, 1369/1990.

STUDIES

Aḥmadi Givi, Ḥ. (2001), *Dastur-e tārikhi-ye feʿl*. 2 vols. Tehrān: Nashr-e Qaṭre, 1380.

Asatrian, G. (1989), *Otglagol'nye imena v srednepersidskom i parfyanskom (na materiale turfanskix tekstov)*. Jerevan: Izdatel'stvo AN armjanskoj CCR.

Bahār, M.-T. (1958), *Sabk-shenāsi yā tārikh-e taṭavvor-e nathr-e fārsi*, 3 vols. Tehrān: Amir Kabir, 1337.

Bausani, A. (2023), *Caratteristiche e sviluppi della sintassi persiana*. Edizione e note a cura di Daniela Meneghini e Paola Orsatti. Pisa: Fabrizio Serra (= *Rivista degli Studi Orientali*, Nuova Serie 96/1. 43-102).

Beelaert, A.L.F.A. (2000), *A Cure for the Grieving. Studies on the Poetry of the 12th-Century Persian Court Poet Khāqānī Širwānī*. Leiden: Nederlands Instituut voor het Nabije Oosten.

Dahlén, A. P. (2011), He addressed the Kayānian king: "I am a prophet" – The Image of Zoroaster in the *Dāstān-e Goshtāsp* (Tale of Goshtāsp). *Orientalia Suecana* 60. 117-39.

Farshidvard, Kh. (1999), *Jomle va tahavvol-e ān dar zabān-e fārsi*. 2nd edition. Tehrān: Amir Kabir, 1378 (1st ed. 1375/1996).

Haspelmath, M. (1995), *The converb as a cross-linguistically valid category*. In M. Haspelmath & E. König (eds.), *Converbs in Cross-Linguistic Perspective. Structure and Meaning of Adverbial Verb Forms – Adverbial Participles, Gerunds –*. Berlin & New York: Mouton de Gruyter. 1-55.

Holland, G. B. (1986), Nominal Sentences and the Origin of Absolute Constructions in Indo-European. *Zeitschrift für vergleichende Sprachforschung* 99/2. 163-93.

Jensen, H. (1931), *Neupersische Grammatik*. Heidelberg: Carl Winter's Universitätsbuchhandlung.

Jespersen, O. (1933), *Essentials of English Grammar*. Boston & Sidney: Allen & Unwin.

Khaleghi Motlagh, Dj. (2001-2009), 4 vols. *Notes on the Shahnameh* (Persian Text Series). New York: Persian Heritage Foundation.

Khāleqi-Motlaq, J. (2002), Negāh-i be hezār beyt-e Daqiqi va sanjesh bā sokhan-e Ferdowsi. In ʿAli Dehbāshi (ed.), *Sokhanhā-ye dirine. Majmuʿe-ye maqāle dar bāre-ye Ferdowsi va Shāhnāme*. Tehrān: Afkār, 1381. 329-406.

Khānlari, P. N. (1986), *Tārikh-e zabān-e fārsi*, 1st revised edition, 3 vols. 1365. Reprint Tehrān: Ferdows, 1377[6]/1998.

Lazard, G. (1963), *La langue des plus anciens monuments de la prose persane*. Paris: Klincksieck, 1963.

Lazard, G. (1964), *Les premiers poètes persans (IXe–Xe siècles): Fragments rassemblés, édités et traduits*. Tome 1: *Introduction et traduction française*. Tome 2: *Textes persans*. Téhéran: Département d'iranologie de l'Institut franco-iranien.

Lazard, G. (2006), *Grammaire du persan contemporain*. Nouvelle édition avec la collaboration de Y. Richard, R. Hechmati & P. Samvelian. Téhéran: Institut Français de Recherche en Iran [1st ed. 1957].

Lenepveu-Hotz, A. (2014), *L'évolution du système verbal persan: Xe- XVIe siècles*. Leuven: Peeters.

Orsatti, P. (2021), *Bausani linguista*. In Claudio Lo Jacono (ed.), *Il retaggio culturale di Alessandro Bausani a un trentennio dalla sua morte*. Roma: Bardi Edizioni. 179-192.

Orsatti, P. (2023), *Alessandro Bausani e la storia della sintassi persiana*, in Bausani 2023. 103-122.

Rubinčik, Ju. A. (2001), *Grammatika sovremennogo persidskogo literaturnogo jazyka*, Vostočnaja literatura RAN, Moskva.

Shafiʿi, M. (1964), *Shāhnāme va dastur*. Tehrān: Nil, 1343.

Shagal, Ksenia. 2019. *Participles. A typological study*. Berlin & Boston: de Gruyter.

Weisser, Ph. (2015), *Derived Coordination. A Minimalist Perspective on Clause Chains, Converbs and Asymmetric Coordination*. Berlin & Boston: de Gruyter.

Windfuhr, Gernot L. 1979. *Persian Grammar. History and State of its Study*. The Hague-Paris-New York: Mouton.

YEREVAN STATE UNIVERSITY PUBLISHING HOUSE

JIL 1 (2024): 66-89

Journal of Iranian Linguistics

About *šekam* and *šotor*: The Development of the Initial Vs/*š*C- in Middle and New Persian*

Salman Aliyari Babolghani
University of Hamburg

doi.org/10.46991/jil/2024.01.04

Abstract: The purpose of this paper is to survey the phonological evolution of the initial sequence Vs/*š*C- in Persian as reflected, for instance, in Middle Persian *iškamb* > New Persian *šekam* 'belly', and to explore philological issues associated with this evolution.

Keywords: Middle Persian, Early New Persian, Southwestern Iranian, Lori, Shirazi-Erahistani, Larestani, Kirmani, Tustari, Caucasian Tati, prothesis, consonant cluster, initial cluster

Salman Aliyari Babolghani
E-mail: aliyari.ba.s@gmail.com

Received: 25.04.2024
Revised: 21.08.2024
Accepted: 02.09.2024

This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License

1. Introduction

In this paper, I aim to examine the phonological evolution of the initial sequence Vs/*š*C- in Persian (Prs.)[1] from the Middle Iranian (MIr.) period onwards. The phonetic context of the development under investigation (formulized here as Vs/*š*C-) is restricted to initial short vowels followed by a cluster comprising sibilants *s* and *š*[2] along with plosive or nasal consonants.

* I am deeply thankful to Prof. Paul and my friends Dr. Pejman Firoozbakhsh and Dr. Meysam Mohammadi for their valuable comments that improved the manuscript, though they may not agree with all of the interpretations and conclusions presented in this paper. I also appreciate the reviewers for their insightful comments, which highlighted points I had not previously discussed.

[1] For abbreviations, see the end of the paper.

[2] Theoretically, the phonetic context of this development could also involve the sibilants *z* and *ž*. However, due to 'Southwestern Iranian' languages (SWIr.) characteristic developments, such as the reduction of *zb* to *z*, the sequence Vz/*ž*C-

It is to be noted in advance that the initial vowel in V*s*/*š*C- can be of two types: (1) a prothetic palatal added to the earlier initial consonant clusters **st*-, **sk*-, **šm*- (< **xšm*-), etc. (here, it is referred to as type a); (2) an inherited original short vowel, or a short vowel derived from earlier initial syllables (such as **abi*-, etc.) or long vowels (here, they are all referred to as type b)[3]. However, since MIr. onwards, both types have converged in a similar phonetic context. Thus, irrespective of their origin, they undergo a shared development from then on[4].

The sequence V*s*/*š*C- eventually yields *s*/*š*VC- in NP, as seen, for instance, in MP ***ušt****ar* 'camel' becoming ***šot****or* in NP. Nonetheless, as will be observed, the treatment of MP and Early NP (ENP), along with the process resulting in the aforementioned transition leads to some ambiguities and discrepancies. These complexities give rise to several debated issues that pose challenges for explanation from the historical linguistics perspective. In the following, first, I will overview the treatment of SWIr. other than Persian regarding the preservation or alteration of this sequence. Afterward, I will return to discussions on the development of this sequence in Persian and the associated issues.

---

may either not exist or occur very rarely in a certain SWIr. One instance of this kind is the word for 'tongue', which appears as *ezbū* in Larestani; *zabân* in (Modern) New Persian (NP) (both < 'Northwestern Iranian' (NWIr.)) vs. *zōn* in Lori (as the true SWIr.). This word could be considered a proper instance for the development in question in Larestani assuming *ezbū* derived from **i****zb****ān*. I am not sure if the same applies to the NP equivalent *zabân* being derived from Middle Persian (MP) *i*/*u****zw****ān* with a different phonetic context. However, one example of this kind in Persian that can be included in our analysis is NP *zomorrod* 'emerald' (cf. § 3.2.1).

For the same reason, i.e. SWIr. characteristic developments, certain clusters of the type in question may hardly ever take place (such as **sk* being changed to *šk*) or be limited to NWIr. loans (such as **sp* being reduced to *s*, cf. below, fn. 4).

[3] The two types of the V*s*/*š*C- have usually been argued in conjunction with the other MIr. (V)CC- such as *fr*-, *dr*-, *afs*/*š*-, etc. (cf. Horn 1898-1901: 39-40; Lazard 1963: 175-176; Ṣâdeqī 1380/2001, esp. 13-16, 20-22; Pisowicz 1985: 127-128; Lenepveu-Hotz 2011), and sometimes, overlooking the fact that either in Persian or other SWIr., they do not show similar treatments and cannot be explained collectively. For instance, unlike the structure under investigation, the obliteration of the Old Iranian (OIr.) initial cluster **dr*- does not occur by adding a prothetic vowel; it is always the insertion of an anaptyctic vowel that breaks the cluster, cf., e.g. NP *dorūγ* (< NWIr.); Lori *dorū* 'lie', etc. A sporadic inconsistent case, however, might be the form <ʾdrm> 'drachm', which Maqdesi reported as existing in the 'language of Bukhara' (see Ṣâdeqī 1380/2001: 14). Nonetheless, the form frequently attested elsewhere in ENP is *dir*(*h*)*am* (cf. also MP *drahm*; NP *der*(*h*)*am*).

[4] Accordingly, NWIr. loans such as MP *ispiš* 'louse' (> NP *šepeš* 'id.' vs. Lori *šeš* 'id.', as a true SWIr. form; cf. Avestan (Av.) *spiš*- 'id.') belong here, being borrowed early enough to be involved in the development.

2. 'Southwestern Iranian' Languages other than Persian[5]

The sequence remains unchanged in all SWIr. other than Persian. In some dialect groups, in particular in Lori, this preservation is highly consistent, whereas in others, some discrepancies arise (discussed after the examples).

Notably, in most of these dialects, MIr. **i*s/šC- yields *e*s/šC-. Moreover, in some cases, the initial short vowel may be lengthened—typically through regressive assimilation affected by a long vowel in the following syllable (cf. below, e.g. ***â***s*âra* 'star'). Nonetheless, the focus here is on the historical significance of preserving the sequence in question or changing it; so, such marginal changes are not under consideration.

Lori[6]

Type (a):

'tent pole' Baxtīârī SL *estīn*; Bâlâgerīva NL *hossīn*, Sagvand NL *hūəs*(*s*)*ī* (< **ustūn* < **istūn*)[7]. Cf. 'column' NP *sotūn*; MP^{Z} <stwn'> read as *stūn*, MP^{M}, Prt^{M} *istūn*; OP <st^{u}una> *stūnā-*; Av. *stūnā-*;

---

[5] This paper is part of a larger research project 'Towards a Historical Dialectology of Lori (Southwest Iran) (DFG-SPP 2176)', initiated in August 2021 under the supervision of Prof. Dr. Ludwig Paul at the University of Hamburg. Through this research, I aim to propose a preliminary classification of SWIr., a hitherto relatively overlooked subject in Iranian philology. Such classification is also reflected in the present paper without detailed argumentations.

[6] Linguistic materials are taken from the sources which are listed here to avoid cluttering the paper with repetitive references. Hereafter, they will be specified only in case of necessity: Achaemenid Elamite (AE) and Achaemenid Babylonian (AB) of the Achaemenid inscriptions from DARIOSH-Louvre Project (in progress); AE of the Persepolis Fortification (PF) from Hallock 1969; Av. from Bartholomae 1904 and Kellens 1995; Old Persian (OP) from Schmitt 2014; Manichaean MP and Parthian (MP^{M} and Prt^{M}, respectively) from Durkin-Meisterernst 2004; Zoroastrian MP (MP^{Z}) from MacKenzie 1990; Inscriptional MP and Parthian (MP^{I} and Prt^{I}, respectively) from Gignoux 1972; ENP from Lazard 1963, Ḥasandūst 1393/2014, and Anvarī 1382/2003; NP examples are from the official NP of Iran; Baxtīārī and Boyeraḥmadi (Beyramey) Southern Lori (SL) from Ṭâherī 1389/2010; 1395/2016, respectively; Sagvand and Dare-J̌owzâni Northern Lori (NL) from Aliyari Babolghani 1396/2017; Shirazi-Erahistani of Fars Province from Salâmī 1383/2004; 1384/2005; 1385/2006; Kumzâri and Lâraki from Anonby and Yousefian 2011, and Lâraki from Aṣγarī 1401/2022; Larestani from Salâmī 1386/2007; 1388/2009; Kirmani of the southern half of Kerman Province, including J̌īroft, Kahnūǰ, Rūdbâr, etc., referred to here as Halīlrūdi, from Nīknafas Dehqânī 1377/1998 and Borjian 2016; Mīnâbi from Barbera 2005; Qešmi from Anonby 2015; Juhuri Caucasian Tati from Authier 2012 and Caucasian Tati of Shirvan (here Širvâni) from Suleymanov 2020. The rest are from the unpublished linguistic materials collected by the present author.

[7] The word in its general meaning, i.e. 'column', takes the form *setīn* (influenced by NP or borrowed from ENP) in most Lori dialects. The true Lori form is, as seen above, preserved in a specific example of 'column', namely, 'tent pole'.

'to break' (pst.) Baxtīârī SL *eška*(*he*)*st-* (int.); Bahmeʾī SL *ešše̊nâ*(*δ*)- (trns.); NL *eškenâ-* (trns.), *eškes-* (int.). Cf. NP *šekast-*, *šekând-*; MP^{Z} <TBLWN-t-> read as *škast-*, MP^{M} *išk̊ast-*; Av. √*sciṇd-* 'to split';

'you' (pl.) SL *īšâ,* Baxtīârī SL *īsâ*. Cf. NP *šomâ*; MP^{M}, Prt^{M} *išmāh*; Av. *xšmākəm*;

Type (b):

'camel' Baxtīârī SL *oštor*, Mamasanī SL *ošter*, also *šotor*; Bâlâgerīva NL *šüter* (< Prs.)[8]. Cf. NP *šotor*; MP^{Z} *uštar* (< NWIr.); OP *ušabāra-* 'camel-borne'; Av. *uštra-*;

'to count' (pst.) SL, NL *ešmārd-*; Sagvand NL *ešmard-*. Cf. NP *šemord-*; MP^{Z} <ʾwšmwlt-> *ōšmurd-* (< **abi-*√*šmar-*); already with a short vowel in MP^{M} *ušmār-* (prs.); *išmīr-* (prs.) (< **ušmīr-*< **abi-*√*šmṛya-*) 'to be reckoned, accounted'[9]; Prt^{M} *išmār* 'number';

'to entrust, consign, etc.' (pst.) Baxtīârī SL and Dare-J̌owzâni NL *espârd-*. Cf. NP *sepord-*; ENP *ispurd-* (apparently via **ō/uspurd-*); MP^{Z} *abespurd-*, *abespārd-*; Prt^{M} *abespurd-*;

Shirazi-Erahistani[10]

Type (a):

'star' Kumzâri *stârg*, but Lâraki *e/istârg*, Behbahâni *âsâra*, etc. Cf. Lori *âsâra*, etc.; NP *setâre*; MP^{Z} <stʾlk'> read as *stārag*, MP^{M} *istārag*; Av. *star-*;

'cave' Davâni *eškat*, Mâsarmi *eškaft*. Cf. NP *šekaft*; MP^{Z} <škpt'> read as *škaft*; MP^{M} *išk̊āft-* 'to split' (pst.);

[8] It should be noted that camels are not commonly raised as domestic animals in Lori-speaking areas, primarily due to the mountainous terrain.

[9] For MP^{M} examples, which do not adhere to Durkin-Meisterernst's (2004: 57, 93) transcription herein, as well as the proposed derivation, see Henning 1933: 193/100, 206/113. Probably also the MP^{Z} equivalent should be read as ***u***šmurdan, ***u***šmār-.

[10] By this term, coined by Dr. P. Firoozbakhsh and me for convenience, I intend the dialect group including the survivals of the former vernacular of the cities Shiraz, Neyrīz, and Kâzerūn, alongside the homogeneous dialects spoken in Fars (usually called 'Tâǰīk(ī)' and more widely 'Fars Dialects') and Bushehr Provinces, as well as Behbahâni and Kumzâri-Lâraki. For details, see Aliyari Babolghani, fort.: appendix.

'belly' former dialect of Shiraz[11], Behbahâni, etc. *ešk*am; Bardestâni *kom* (< **išk*amb, with the omission of the first syllable); Kumzâri *šk*om, but Lâraki *ešk*om[12]. Cf. Lori *ešk*am[13]; NP *šek*am; MP^{M} *išk*amb[14];

consider also Davâni, Bardestâni, etc. *ešk*a:s-, Behbahâni *ešk*ess- 'to break' (pst., int.); several dialects *šūmū* or *šomâ*, but Kumzâri *šmâ* and Lâraki *ešmâ* 'you' (pl.).

Type (b): Several dialects *šotor*, but Dūsīrâni *oštor* 'camel'; Davâni *ešmord-*, Mehbūdi *ešmârd-*; Lâraki *ešma:rd-*, Kumzâri (*e*)*šmârd-* 'to count' (pst.); Dūsīrâni, Davâni, etc. *espord-* 'to entrust, etc.' (pst.).

Larestani[15]

Type (a): Evazi, Gerâši, etc. *ešk*at 'cave'; Evazi *aškom*, Xonǰi *oškom*; Asīri, Aheli *kom* (< **išk*amb) 'belly'; Xonǰi *ešk*ehes- (int.), Asīri *ešk*ahond- (trns.), *ešk*at- (int.) 'to break' (pst.), etc.; Xonǰi *essara*, Fīšvari, Evazi *estara*, etc. 'star'; several dialects *šomâ*, but Gerâši *īšnīâ* 'you' (pl.).

Type (b): Aheli, Xonǰi, etc. *ezbū* 'tongue' (cf. NP *zabân* 'id.'; MP^{Z} *uzwān*, *zuwān*, MP^{M} *izwān* 'id.' (< NWIr.)[16]; Prt^{M} *izβān* 'id.'; OP *hizānam* 'id.'; Av. *hizuuā-* 'id.'); Aheli, Xonǰi, etc. *oštor* 'camel';

Kirmani[17]

Type (a): North Baškardi (NB), Halīlrūdi *estâl*, Qešmi *estâla* 'star'; Halīlrūdi *ešk*am, Mīnâbi *e/oškom*, Qešmi *ešk*om 'belly'[18]; Mīnâbi *ešk*aht- 'to break'

[11] See Firoozbakhsh 2019: 181, 183, ghazal 44, line 4.

[12] Lâraki *ešk*om, *e/istârg*, and *ešma:rd-* quoted in this section are derived from a personal interview with a Lâraki informant.

[13] The words generally used for 'belly; stomach' in Lori include *kom* (Baxtīârī SL also *ešk*am) in SL and *gīa*, *gada*, etc. in NL (also Baxtīârī SL *gaδe* 'stomach'). The form *ešk*am (cf. *kom*) is used with slightly different meanings or in specific contexts, such as NL *ešk*am-*eš poř bī* 'she was pregnant (lit. her belly/ womb was full') or Bâlâgerīva NL *mīn-ešk*am 'abdominal organs'.

[14] The etymology of the word may be a subject of debate, but there is no dispute regarding the inclusion of an earlier **s/šk-* in its root, cf., e.g. Korn 2005: 349; Cheung 2007: 344-345, and derivations quoted in Ḥasandūst 1393/2014: 1886-1888.

[15] Also known as 'Ačomī' (< Larestani *a-č-om* 'I go'), spoken in Lârestân County, in south Fars Province, as well as the western half of Hormozgan Province. For details, see Aliyari Babolghani, fort.: appendix.

[16] Whereas Lori *zō*(*n*), *zõw*, etc. should go back to the true SWIr. **hizān*(*a*)-. Cf. also fn. 2.

[17] By the term Kirmani, I intend Baškardi and the homogenous dialects in other regions of Hormoz Province as well as the southern half of Kerman Province.

[18] NB has *lâv/w* (cf. Balochi *lâp*) with a distinct origin.

(pst.) (cf. also Halīlrūdi *ešk*a*nd* 'a break or outflow point in a stream'); NB *ešk*o*wt* 'cave'; NB *espīr* 'white' (cf. NP *sefīd* 'id.'; MP^{M}, Prt^{M} *ispēd* 'id.'); Halīlrūdi *espore* 'shovel footpad' (cf. Dare-J̌owzâni NL *espâra* 'id.'; MP^{Z} *ōspurdan*[19] 'to tread, trample').

Type (b): South and NB *ešter*, Mīnâbi *e/oštor*, Qešmi *eštor* 'camel'; NB *ešmârt-*, Mīnâbi, Qešmi *ešmord-* 'to count' (pst.).

Tustari[20]

Type (a):

'to take' (pst.) Šūš. *esad-*, Dez. *osond-*. Cf. SL *es*[*t*]*ey*(*δ*)-, NL *ēsa-*, etc.; NP *setând-*; MP^{Z} <YNSBWN-t-> read as *stad-*, MP^{M}, Prt^{M} *istad-*, from *√*stan-* 'to take (away)'[21];

'ember' Dez. *ežgel*. Cf. Baxtīârī SL *azgel*; NL *ezgel*, *ežgel*; NP *zoγâl* 'coal'; ENP *zugāl*, *sukār*, *sikār*(*a*), *aškar*, *uškār*[22]; Sogdian <sqʾr>, <ʾskʾr> 'coal'[23]; Khotanese *skara-* 'id'[24];

consider also *eška:s-* 'to break' (pst.); *âsâra* 'star'; *eškam* 'belly'.

Type (b): *eštow* 'haste, acceleration' (cf. Baxtīârī SL *eštaw* 'id.'; NP *šetâb* 'id.'; ENP *šitāb*, *i/uštāb* 'id.'; MP$^{M, Z}$ *awištāb* 'oppression' < *awištāb-* 'to oppress; hasten'); Šūš. *ešmârd-* 'to count' (pst.).

Caucasian Tati

Apart from some inconsistent paradigms (see below) such as Juhuri *šumorde* 'to count' (the sole example of type b that I could find in materials at my disposal) the same treatment is seen in the Caucasian Tati as well:

---

[19] Or rather *uspurdan*, cf. below, § 3.2.1.

[20] By this term, I refer to the dialects spoken in the cities of Šūštar (Šūš.) and Dezfūl (Dez.) in Khuzestan Province.

[21] See Henning 1933: 189/96.

[22] For the latter three forms, see Ravâqī 1381/2002: 25, 227. It seems, according to derivations cited in Ḥasandūst (1393/2014: 1567, 1746), that *sikār*(*a*), etc. and *zuγâl* had not been connected before.

[23] Gharib 1995: 61a, 354a.

[24] Bailey 1979: 429. The word's derivation is obscure (for some of the propositions, see Ḥasandūst 1393/2014: 1567, 1746) and consequently, its attribution to neither of types (a) and (b) is certain. It is hypothetically classified here, considering that Bailey (ibid.) links Khotanese *skara-* to Av. *ātrəm skairyaṯ hačā* 'fire from charcoals', etc., and Morgenstierne (2003: 74) derives the Pashto equivalent *skor* 'coal' from **skāra-*. It is also uncertain whether the word is genuine or borrowed in Persian, Tustari, and Lori. Consider that some SL have a distinct word for 'ember', cf., e.g. Boyeraḥmadi and Mamasanī SL *xərong* (cf. MP^{Z} *x^{w}arg*). Nonetheless, the word is an example of the sound change (cf. fn. 2 and 4).

Type (a): Širvâni *ustoran* 'to get; buy'; *ɨškam* (also *šɨqam*) 'belly'; *ispiɦ* (also *sibiɦ*) 'white'; *ɨškɨn* (also *šɨqɨn*) 'landslide' (cf. NP *šekan-* 'to break' (prs.)); Juhuri *išmū*, Širvâni *išmun* 'you' (pl.); Juhuri *astare* 'star'.

Most of the discrepant paradigms occurring in these dialect groups align with the Persian structure of the sequence. Cases such as *setâra/e* 'star', *šotor* 'camel', or even *šomâ/šūmū* 'you' (pl.) in several Shirazi-Erahistani and Larestani dialects, and *sotūn* 'column', *šekâl/r* 'prey', and possibly even *šomah*, *šemâ*, etc. 'you' (pl.) in Kirmani, as well as *šemâ/ō* or *šomâ*(*n*) in NL, fall into this category, likely under the influence of Persian. This should also apply to some similar paradigms in Caucasian Tati such as Širvâni *sibiɦ*, Juhuri *sipi* 'white' (cf. ENP *sipēd* 'id.') and Širvâni *sūtūn* 'column' (besides ENP *sutūn*, cf. Azerbaijani Turkish *sütun* < Prs.). However, in the case of Caucasian Tati, the influence of Turkish might also be considered.

A second type of discrepancy is forms with *s/š*C- frequently observed in Kumzâri. This should be understood as the outcome of a secondary and relatively recent change, namely the apheresis of V*s/š*C-, rather than, for example, the preservation of OIr. **s/š*C-, as one might speculate. This becomes particularly evident when comparing these forms to the equivalents with V*s/š*C- in Lâraki, the more conservative variety of the same idiom.

3. Persian

3.1. Challenges and Current Explanations

In Persian, we observe a markedly different treatment compared to other SWIr. What is clear is the eventual contrast between Persian *š**e**kam*, *š**o**tor* vs. ***e**škam* and ***o**štor*, and so on in other SWIr. However, there are still several ambiguous and disputable aspects regarding this development in Persian that warrant further discussion, as outlined below:

(1) the starting date and the process of such development in Persian;

(2) the issue of the distinct spellings in MP^{M} and MP^{Z}, viz. the fact that the continuations of the OP words with the initial consonant clusters *s/š*C- (type a) are written with a prothetic vowel *i-* (represented by the letter ayin <ʿ->, and less frequently alif <ʾ->) in MP^{M} and without it in MP^{Z} (e.g. MP^{M} <ʿstʾrg> vs. MP^{Z} <stʾlk'> 'star');

(3) the presence of 'dual spellings', i.e. written with and without a word-initial alif, for both lexicons type (a) and (b) in ENP (e.g. <ʾstʾrh> ~ <stʾrh> 'star'; <ʾštr> ~ <štr> 'camel').

Whether explicitly stated or not, the second issue is presently understood as a dialectal variation in MP. Specifically, OP *s/šC-* is preserved as such in MP^{Z} whereas taking a prothetic palatal vowel and changing into *is/šC-* in MP^{M}[25]. However, this distinction disappears in the Early New Iranian (NIr.) period, when Persian is, alongside the Manichaean script, written in two new scripts: Arabo-Persian and Hebrew. ENP texts—irrespective of the script, thus including Manichaean ENP (ENP^{M}) and Early Judaeo-Persian (ENP^{J}) too—surprisingly feature forms both with prothetic and anaptyctic *i*; evidence of such forms can be found even simultaneously in the same text and even in the same manuscript[26]. In Lazard's words: "les deux types de formes alternent dans nos textes, sans qu'il soit possible de trouver un principe à la répartition"[27]. Eventually, in NP, forms with anaptyctic *i* (later > *e*) become dominant in type (a) words, and similarly, forms with an anaptyctic vowel in type (b) words, as seen in the following examples:

Type (a): ENP *istaδ- ~ sitaδ-* (cf. NP *setând-*) 'to take' (pst.); *istāra ~ sitāra* (> NP *setâre*) 'star'; *išk*am *~ šikam* (> NP *šekam*) 'belly'; *išk*ast- *~ šikast-* (> NP *šekast-*) 'to break' (pst.);

type (b): ENP *ušmār ~ šumār* (> NP *šomâr*) 'calculation'; *uštur ~ šutur* (> NP *šoto*r) 'camel'; *ispurda ~ sipurda* (> NP *seporde*) 'delivered' (cf. MP^{Z} *abespurd-*, cited above).

Both spellings are already found in the **earliest attestations of ENP** as well:

(1) *ušnuhil* 'gratitude' (cf. MP^{M} *išnōhr* 'id.'; Av. *xšnaoθra-* 'satisfaction'), found in a translation of *Fātiḥa* (the opening Surah of the Qur'an), probably from the early 9th century or before[28];

---

[25] Cf., for instance, Ṣâdeqī 1380/2001, esp. 15-18; Paul 2013: 53; Rezai Baghbidi 2017: esp. 88; and above all, MacKenzie's (1990) transcription system for MP^{Z}, which is widely accepted by scholars.

[26] For ENP and ENP^{J} examples, see Lazard 1963: 175-176 and Paul 2013: 53-54, respectively. Regarding ENP^{M}, cf. <šnʾsydˀ> ~ <ʿšnʾsyd> 'he recognizes' in the same text (see Sundermann 2003: 256: b16, 257: c3). Given that the scribes of the ENP^{M} texts were generally inclined towards maintaining historical (i.e. MP) spellings (see Henning 1962: 89-90; Sundermann 2003: 245; de Blois 2006: 93-96, and cf., e.g. <ʿstʾr**g**> 'star', as a clear instance belonging here), one might read cases such as ENP^{M} <ʿšnʾsyd> (mentioned above), <ʿspʾḫ> 'army', <ʿstbryḫ> 'harshness', etc. exclusively with the anaptyctic *i*, i.e. *šināsaδ*, *sipāh* and *sitabrī* (as in de Blois 2006: 100). However, compared to the same dual spellings attested elsewhere in the ENP text, the variant forms with prothetic *i-* should have, at least for some words, existed too.

[27] Lazard 1963: 175.

[28] First published by Zadeh (2015, see esp. pp. 402-403). This translation is attributed to Salmān al-Fārisī, the Iranian companion of the Prophet Muhammad. However, the text is documented in the 11th century and its attribution to Salman is questioned. Nonetheless, it is undoubtedly the oldest translation of the Qur'an,

(2) *ispās* 'gratitude' (cf. MP^{M} *ispās* 'id.'; NP *sepâs* 'thanks', and esp. ENP^{J} ***i****pās*, mentioned below) following the quoted translation in the very text for explaining *ušnuhil*;

(3) *iškam*, *iškamb* ~ *šikanb* 'belly' in Persian quotations from the era of Muhammad attested in Arabic texts from the 9th century[29];

(4) <šmr> *šumār* 'reckoning'[30] as well as <sbʾs> *sipās* 'service, thanks'[31] (cf. *ispās*, quoted above), attested in two letters written in Judaeo-Persian, known as Dandān-Uiliq letters no. 1 and 2, dated to the mid-8th and the early 9th century, respectively[32].

(5) <ʾṣṭxr> *Iṣṭaxr* '(the mint of) Istakhr' on Umayyad (661-750 CE) *dirham*s[33], cf. MP^{I} <stḥly, stʾḥly>, MP^{Z} <stʾhl> read as *Staxr*[34], presumably from OP **staxra-* 'strong(hold)'[35]. However, this evidence involves a proper name occurring not in a Persian but in an Arabic text. Therefore, one might consider it inconsistent with other instances mentioned here, interpreting the prothesis as an Arabic adaptation (i.e. *Staxr* pronounced as *Iṣṭaxr* in Arabic) rather than as a reflection of *Istaxr* in its Persian origin. On the contrary, I believe this pronunciation was already present during that period of ENP. Notably, the same form <ʾṣṭxr> *Iṣṭaxr* is frequently attested in later ENP texts, alongside the less common forms <ṣ/sṭxr> *Sitaxr* and <sṭrx> *Sitarx*, found, for instance, in Ferdowsi's *Shahname*[36].

Consider also the fact that already in Ferdowsi's *Shahname* (written in the late 10th century), as an instance, the forms with the anaptyxis, such as *sipahbad* 'general' (with hundreds of attestations. Cf. MP^{M} *ispāh* 'army'), occur with significantly higher frequency than those with the prothesis, such as *ispahbad* 'id.' (with 12 attestations)[37]. However, the latter forms seem to persist until the end of the ENP.

---

dating to around 200 Hijri (ca. the early 9th century) or earlier, and probably originating in Basrah (see Firoozbakhsh 2024).

[29] See Ṣâdeqī 1357/1978: 61, 64.

[30] In Du^{1} 21 and Du^{2} 19 (see Utas 1968: 128-130; Zhang and Shi 2008: 83-86, 94, respectively).

[31] In Du^{2} 25 (see Zhang and Shi 2008: 83-86) Notably, readings *šmar* and *spās* for <šmr> and <sbʾs> (Zhang 2023: 109-111, 113-115, 127, 129) are not acceptable. Cf. MP forms of <šmr> (such as MP^{M} *ušmār-*, with an original initial vowel), cited above in § 2.

[32] Cf. Paul 2013: 10 and references.

[33] See Walker 1956: lxxii.

[34] For instance, in *Ardā Wirāz Nāmag* 1: 5 (see Gignoux 1984: 36, 37, 265).

[35] See Bivar and Boyce 1998.

[36] See Xâleqī-Moṭlaq 1398/2019: 80.

[37] Cf. Xâleqī-Moṭlaq 1398/2019: 154, 237-238. For some further instances, see ibid. 32, 55, 79-80, 152-155, 237-238, 262-265, 267, 270-274, 302, 342-344, 350-353, 430-432, 503-504.

The dual spellings uncategorizably attested in ENP raise the question of what happened to the supposed dialectal variation and how the simultaneous occurrences of these two spellings can be explained. In the case of words like *uštur*, etc. (categorized here as type b), Ṣâdeqī posits that the development into the form *šutur*, etc. did not take place through the shift of the prothetic vowel to an anaptyctic one. He asserts that in the first place, the initial vowel dropped (or changed into *ə*), and then the resulting consonant cluster split by inserting an anaptyctic *ə* which would later change into *a*/*i*/*u* depending on the phonetic context. Given that his argument primarily relies on MP^{Z} forms, it seems that he also considers the insertion of an anaptyctic *ə* applicable to the type (a)[38]. Thus, he regards contrasting paradigms such as *sipōxtan*, *šikōfa* as "exceptions"[39]. This would paradoxically imply that Persian tended to reproduce new initial consonant clusters during the period when it actively avoided them[40]—a point that the author himself alludes to[41].

3.2. A Proposal

I believe we are facing obstacles in reaching a reasonable and commonly acceptable explanation for such forms in ENP because our current arguments are based on an incorrect supposition regarding the prior development of examples of type (a) (cf. issue no. 2, mentioned in § 3.1). I suppose *s*/*š*C- > *is*/*š*C- occurred in early times (at the latest in Early MP (EMP)) and served as a universal, rather than dialectal, sound change in Persian, although in MP^{Z}, it was veiled beneath the cover of the Pahlavi script. In other words, MP^{Z} underwent the same development, thus inherited the same forms as attested in MP^{M}, and featured, e.g. *istārag* and *išk*a*st* rather than *stārag* and *šk*a*st*.

This is a common development in all SWIr. (cf. above) up to this phase. Hereafter, Persian commits the innovation of shifting the vowel of the structure V*s*/*š*C- (in both types a and b) from the beginning into the middle of the cluster. Then, naturally, this vowel could later undergo secondary changes depending on the phonetic environment, especially the quality of the vowel of the following syllable. In many cases, either before or after the vowel shift, vowels *u* and *a* were probably inclined to turn into *i*, due to analogy with the high number of paradigms featuring *is*/*š*C in ENP and *s*/*ši*C- in (E)NP, cf., e.g. ENP *s**i**tān* 'laying on the back' < MP^{Z} ***u**stān*˚ '[with] outspread/outstretched [hands (in prayer)]'; NP *s**e**tordan* 'to erase, shave' < ENP

[38] The position that the author takes here is not precisely clear to me.
[39] Ṣâdeqī 1380/2001, esp. 15-18. Cf. also Pisowicz 1985: 127-128, 146-147.
[40] Cf. also Lenepveu-Hotz 2011: 84-86.
[41] Ṣâdeqī 1380/2001: 22.

u*sturdan* 'id.'; NP *šetâb* 'haste' < ENP *ištāb* < *uštāb* < MP *awištāb* 'oppression'. From ENP onwards, the older forms (in my view) with V*s*/*š*C- gradually fade away in favor of those with *s*/*š*VC-, until eventually in NP, the latter forms become quite dominant. The reasons and pieces of evidence that led me to such an assumption are as follows:

(1) The addition of a prothetic *i*- (> *e*-) to the initial consonant clusters under investigation (i.e. type a) is a universal treatment in SWIr. which is widely observed also in NWIr., with Middle Prt. being attested earlier (cf. the Prt^{M} equivalents such as *istūn* 'column', *išmār* 'number', etc., cited so far). This fact would per se indicate the antiquity of the evolution. On the other hand, MP^{M} clearly shows that Persian had also undergone the same change, so it would be surprising if MP^{Z} had exceptionally resisted such a common and relatively old development.

(2) Generally, MP^{M} attests to more conservative forms, while MP^{Z} contains more innovative ones. It would be unexpected for MP^{Z} here to conservatively preserve the earlier *s*/*š*C-.

(3) The development occurring in consonant clusters of type (b), as in MP *uštar* > (E)NP *šutur*, suggests that type (a) should have undergone a similar process—i.e. the shift of the prothetic vowel to an anaptyctic one, e.g. MP *istārag* > (E)NP *sitāra*. It is not accidental that the dominant anaptyctic vowel here in type (a) is *i* (> *e*). Hence, unlike what Ṣâdeqī[42] suggests, cases like (E)NP *šikōfa* are not "exceptional", but according to the rule.

(4) If such a dialectal distinction ever existed in MP, the same distinction should have been reflected in some ENP texts, whereas we consistently encounter a mixture of the two spellings in all ENP texts. My interpretation is that MP^{M}-type forms with the prothetic *i*- are continued up to ENP. However, being in the course of development, these forms are attested simultaneously and closely associated with the innovative forms featuring the anaptyctic *i* (e.g. ***i****stāra* ~ *s****i****tāra*, etc.) until eventually in NP, the latter forms (i.e. *s****i****tāra* > *s****e****târe*, etc.) become dominant. The sequence of this development, i.e. MP *is*/*š*C- > ENP *is*/*š*C-/ *s*/*ši*C- > NP *s*/*ši*C-/ *s*/*še*C-, per se contradicts the assumption of the preservation of OP *s*/*š*C- in any MP dialect.

(5) The main obstacle against my supposition is that such pronunciation is not reflected in the Pahlavi script. An adequate explanation can be obtained only through a separate investigation. However, as far as our subject is concerned, it can be asserted that even though the earlier **s*/*š*C- is written with <s/šC-> sign sequences (without the prothetic vowel, as claimed here) in the Pahlavi script, under certain conditions, evidence of the

[42] Ṣâdeqī 1380/2001: 18.

prothesis in question can be found in this script too, which is discussed in the successive section.

3.2.1. Reflection of the Prothetic Vowel in the Pahlavi Script

Middle Persian *is/šC-* (< OIr. **s/šC-*) > MP^{Z} *as/šC-*, *us/šC-*

The first condition leading to the emergence of the prothetic vowel of the <*s/šC*-> words in the Pahlavi script arises when the prothetic *i*- in *is/šC*- (type a), through a secondary change, had the chance to transform into other short vowels, resulting in *us/šC*- or *as/šC*-:

(1) MP^{Z} <ʾškmb'> *aškamb* 'belly, womb', via regressive assimilation, from *iškamb*, the earlier form that is attested in both MP^{M} and Prt^{M}, cf. also ENP *iškam* and the equivalents in other SWIr. mentioned earlier;

(2) MP^{Z} <spwlt-n', spl-> read as *spurdan*, *spar*-; <ʾwspl-tn'> read as *ōspurdan*, *ōspar*-; <wspwl-tn'> read as *wispurdan*, **wispar*-, all conveying the same meaning of 'to tread, trample'. However, I propose that these variations are likely only graphic, all essentially representing *uspurdan*, *uspar*-[43] which later gives ENP *ispurdan*, *sipurdan*;

(3) MP^{Z} <ʾwšnwk'> *ušnūg* beside the spelling <šnwk'> read as *šnūg* 'knee'. Cf. MP^{M} <ʿšnwg> *išnūg*; Av. (*x*)*šnu*-;

(4) MP^{Z} <ʾspnc> *aspiǰ* 'hospitality; inn' beside the spelling <spncʾnkyh> read as *spinǰānagīh* 'hospitality'. Cf. ENP (*sarā i*) *sipanǰ* 'inn'; MP^{M}, Prt^{M} <ʿspync/j> *ispenj* 'id.'[44];

(5) MP^{Z} *ašmā*, to the best of my knowledge, is exclusively written in the *huzwāreš* <LKWM>. However, if we accept the current reading, it could serve as indirect evidence relevant to this section. Cf. MP^{M}, Prt^{M} *išmāh*[45]; Av. *xšmākəm*;

---

[43] Consider that /*u*-/ in the Pahlavi script can be represented by <ʾw->, as seen in, e.g., <ʾwštl> *uštar* 'camel' and <ʾwspwlyk'> *uspurrīg* 'complete'.

[44] This word could belong here, but it is uncertain due to the ambiguity in its derivation (some of them quoted in Ḥasandūst 1393/2014: 1676). Henning states that *aspiǰ* "may be a derivative of MPers. *asp*- (Man. *hasp*-) 'to rest', *aspīn* (Man. *hspyn*) sbst. 'rest' […], so that *sipanǰ* would mean 'rest-house' even by etymology" (Henning 1965: 244/619: fn. 11). If this is the case, this example should be disregarded here. However, the mentioned derivation encounters some phonological obstacles which are left unexplained. Indeed, the Pahlavi spelling with <spˀ> corresponding to that of the MP^{M} and Prt^{M} with <ʿspˀ> would probably suggest that its OIr. origin started with **sp*-.

[45] Prt^{M} <ʿšmʾ(ḥ), ʾšmʾḥ> read as ***i****šmāh*, wherease MP^{M} <ʾšmʾ(ḥ/h), ʿšmʾ(ḥ), etc.> as ***a****šmāh* by Durkin-Meisterernst (2004: 56, 92). Not only in MP^{M}, but probably also in MP^{Z} the pronunciation should likely have been *išmā*(*h*) (as in Prt^{M}) rather than *ašmā*(*h*). This is also supported by the spellings in ENP^{J} <yšmʾ> *išmā* (in Du^{2} 7, see Zhang and Shi 2008: 82-83, 85-86; cf. Paul 2013: 95-96, 100) and ENP <ʾšmʾ> ~ <ʾyšmʾ> *išmā* (in *Tafsīr-i Sūrābādī*, see Ravâqī 1381/2002: 25, 38).

(6) MP^{Z} <ʾwzmbwlt'> read as *uzumburd* 'emerald', borrowed from Greek *smáragdos* 'id.'. This word can also be included here as an example of the similar phonetic context *z*C- (cf. fn. 2 and 4), specifically *zm-* < *sm-*, where *s* became voiced before *m*. Cf. also Armn. *zmrouxt* 'id.' (< Iranian)[46]; NP *zomorrod* 'id.'.

Middle Persian privative prefix *an-*

Another context in which the prothetic vowel appears is in the combination of <*s*/*š*C-> Pahlavi words with the MP privative prefix where the prefix is occasionally written in its prevocalic variant, namely <ʾn-> *an-*[47], cf., e.g. (1) <ʾnspʾs> *an-ispās* (beside <ʾspʾs> read as *a-spās*) 'ungrateful' and <ʾnspʾsyh> *an-ispāsīh*[48] 'ingratitude'; (2) <ʾnšnʾsk'> *an-išnāsag* 'unknown, unidentifiable'; (3) <ʾnšnwhlyh'> *an-išnōhrīhā* 'having no gratitude (to gods)'[49].

Pahlavi <s/šC-> rendering original Vs/*š*C-

Furthermore, a handful of words of type (b) may, in a distinct manner, indicate a similar orthographical behavior. In the following examples, we encounter Vs/*š*C- with an original initial vowel, which remains unrepresented in the Pahlavi script:

(1) MP^{Z} <spwlyk'> read as *spurrīg*, beside <ʾwspwlyk'> *uspurrīg* 'complete', derived from **us-√parH-* 'to fill'[50]. Cf. MP^{M}, Prt^{M} <ʿspwr> *ispurr* and <ʿspwryg> *ispurrīg* 'id.'; (E)NP *siparī* 'complete, ended, etc.'. Additionally, consider MP^{Z} <ʾ**n**wspwlyk'> and <ʾ**n**spwl>[51] 'imperfect', which could respectively represent *an-uspurrīg* and *an-ispurr* (cf. below), the latter reflecting the more recent pronunciation.

(2) MP^{Z} <stwb'> read as *stō* 'distressed, defeated', derived from **us-√tav-* 'to be able'[52]. Cf. ENP *u*/*istōh*, *sutōh*; MP^{M} <ʿstwy-> *istōy-* 'to defeat'; <ʿstwyqwn> *istōy-kun* 'conqueror'; Prt^{M} <ʿstwb-> *istōβ-* 'to defeat'; <ʿstwb> *istōβ* 'defeated'.

---

[46] See Schmitt and Bailey 1986.
[47] I am grateful to my friend Dr. Yusef Saadat for bringing this to my attention.
[48] MacKenzie (1990: 10) reads them as *an-espās* and *an-espāsīh*, respectively.
[49] The two latter attested in *Dēnkard V* 15: 5 and 24: 21, respectively (see Amouzgar and Tafazzoli 2000: 54, 55, 94, 95, 130). Amouzgar and Tafazzoli (ibid.) read them as *ana-šnāsag* and *ana-šnōhrīhā*, respectively.
[50] See Cheung 2007: 295-296 and references.
[51] In *Dādestān ī Dēnīg* 36: 2 (see Jaafari-Dehaghi 1998: 108, 242).
[52] See Ghilain 1939: 67. Cheung (2007: 367) criticizes this derivation, and proposes a new one assuming the root **√staup-* 'to overcome, defeat', based solely on the abovementioned cases.

One might simply explain these spellings by assuming the deletion of the initial vowel. However, the presence of the initial *i-* in the MP^{M}, Prt^{M}, and ENP equivalents contradicts such an assumption. Instead, it suggests that MP^{Z} <spwlyk'> and <stwb'> were likely pronounced with *is*C- (< *us*C, as occurs in the MP^{M}), i.e. *ispurrīg* and *istō*, respectively. If so, they would, from another perspective, lend support to the previously mentioned assumption suggesting that the Pahlavi script may avoid reflecting the first vowel of V*s*/*š*C-, when that vowel is *i-*.

3.2.2. Reflection of the Anaptyctic Vowel in Pahlavi

ENP forms with anaptyctic *i* (such as *š**i**kanb*, cf. § 3.1) are already attested in the early centuries after Islam. Thus, it can be theoretically assumed that the forward shift of the prothetic vowel (e.g. ***i**stāra* > *s**i**tāra*), might have begun before that time, namely, in the Late MIr. period. There is at least one instance that supports this assumption.

In a paronomasia found in the Pahlavi text *Andarz ī Pōryōtkēšān*, the word <spʾsdʾl> 'grateful' is interpreted through folk etymology as *sē*/*i-pās-dār* 'one who keeps three watches'[53]. This example documents the pronunciation *sipās-dār*, suggesting that the vowel shift had already commenced during the Late MIr. period. It also indicates that, in late Pahlavi texts, some words of type (a) (written with <s/šC->) may have already been pronounced with an anaptyctic vowel.

3.2.3. Old Persian Initial <s/šC-> in Achaemenid Elamite Garb

As previously mentioned, I posit that the addition of the prothetic *i-* likely occurred by EMP. However, it can be hypothesized that this phenomenon dates back to earlier periods, possibly to that of Old Persian (OP). In Achaemenid Elamite (AE) renderings of OP words, the clusters under investigation are consistently represented by the *iš*-CV—more specifically *iš*-CV(C)—sign sequences. The same pattern, although it is less regular, is observed in Achaemenid Babylonian (AB) cuneiform. Consider the examples:

[53] The text reads: *mardōm kē-š ēn sē*/*i pās ī-m guft abar tan ī x^{w}ēš bē pāyīd…, ēg sipās-dār*/*se-pās-dār būd, ud pad sipās-dārīh*/*se-pās-dārīh ēn tuwān kardan kū ruwān ō dušox nē rasēd* 'people who keep these three watches, which I mentioned, on their own body… they shall become 'grateful' ('one who keeps the three watches'), and through 'gratitude' ('keeping the three watches'), one shall be able <to avoid> reaching hell' (for details, see Qâʾemmaqâmī 1401/2022: 402-405, esp. 404: fn. 1). The transcription and translation of the passage are based on Qâʾemmaqâmī's reading rather than being a direct quotation.

(1) OP <sk^{u}udr> 'Thracia; Thracian': AE ^{DIŠ}*iš-ku-ud-ra*, AB $^{KUR/LÚ}$*is-ku-du-ru*(-$^{\cdot}$)[54];

(2) OP <sprd>[55] 'Lydia': AE $^{DIŠ/AŠ}$*iš-pár-da*, but AB ^{KUR}*sa-par-da*/ ^{KUR}*sa-pa-ar-da*[56];

(3) OP [<st^{u}una>] 'column': AE ^{AŠ}*iš-du-na-um*[57];

(4) OP <stanm> 'place': AE ^{AŠ}*iš-da-na*[58];

(5) OP <sk^{u}ux> personal name: AE ^{DIŠ}*iš-ku-in-ka*$_{4}$[59].

The OP cuneiform itself never reflects *is*/*šC*- < **s*/*šC*-, making us believe that it is merely an orthographical convention in AE cuneiform for rendering OP *s*/*šC*-; so *i*- here is only graphic. However, this matter may not be established so straightforwardly. If AE *iš*-CV, as a VC_1-C_2V cuneiform sign sequence type, was employed for rendering OP *s*/*šC*-, theoretically, other sign sequences of this type should have had an equal chance of being utilized for the same purpose. We are aware that *uš*-CV was impractical since the sign *uš* was already out of use in AE but *áš*-CV was expected to be regularly documented, resulting in spellings like AE ^{*DIŠ}*áš-pár-da* as a variant of ^{DIŠ}*iš-pár-da* ~ OP <sprd>, and so on. However, such variant spellings do not occur in AE.

Furthermore, employing the AE VC_1-C_2V type of sign sequence—one example of which is *iš*-CV—is not the habitual method of Elamite scribes for representing OP initial consonant clusters, cf. e.g. AE *pír-r*V, of the type C_1VC_2-C_2V, representing OP *fr*- and *br*-, for instance in ^{DIŠ}*pír-ra-da* ~ OP <frad> *Frāda* and AE *pír-ra-iz-man-nu-ia* ~ OP <brzmniy> *brazmaniya*[60]. AE *iš*-CV, in fact, echoes AE *ir*-CV(C) sign sequences systematically used for rendering OP *ṛ*C-, as seen in, e.g. AE ^{DIŠ}*ir-tak-ik-šá-áš-šá* ~ OP <artxšça> *Ṛtaxšaça*-[61].

Accordingly, I suppose *i*- in the AE *iš*-CV should indicate a linguistic fact rather than being purely graphic. Two possibilities could be hypothesized: (1) it reflects the Elamite phonological adaptation of OP initial clusters of this kind. For instance, Elamite-speakers may have pronounced OP *stānam* as

---

[54] In DNa^{OP} 29/ DNa^{AE} 23-24/ DNa^{AB} 17; A^{3}Pb^{OP} 25/ A^{3}Pb^{AE} 25 (here ^{DIŠ}*iš-ku-ra*)/ A^{3}Pb^{AB} 25. Also in PF, e.g. AE ^{DIŠ}*iš-ku-tur-raš* (PF 1820: 4-5; PF 1823: 4-5), AE ^{DIŠ}*iš-ku-ud-ra-ip* (PF 1056: 3; PF 1085: 3).

[55] From Lydian *Śfarda*-.

[56] In DNa^{OP} 28/ DNa^{AE} 22/ DNa^{AB} 16; DHa^{OP} 6/ DHa^{AE} 5-6/ DHa^{AB} 6; XPh^{OP} 22/ XPh^{AE} 18/ XPh^{AB} 18). Also in PF, e.g. AE ^{AŠ}*iš-pár-da* (PF 1321: 8-9; PF 1404: 7-8, etc.).

[57] In DSz^{OP} y+5/ DSz^{AE} 42.

[58] In XVa^{OP} 20-21/ XVa^{AE} 20-21.

[59] In DBk^{OP} 1-2/ DBk^{AE} 1.

[60] For further examples, see Mayrhofer 1973: 41-42, 64, 67.

[61] For further examples, see Mayrhofer (1973: 25), and cf. R. Schmitt's transcription system for OP.

**is/štanam* or the like, and so on; (2) it testifies a phonological aspect of OP, i.e. earlier **s/š*C- > *is/š*C- or *əs/š*C-, not reflected in the OP script itself[62].

However, unlike the latter assumption, the comparable OP word *ištiš* 'brick' (cf. Av. *ištiia-* 'id.') is spelled as <ištiš> with *i-*. One hypothetical explanation might be that the words under discussion were pronounced differently, viz. as *əs/š*C- rather than *is/š*C-. Alternatively, the presence of *i-* in <ištiš> might be due to its pronunciation as **hištiš*, with the prothetic *h-* dating back to OP (cf. Prt^{M} *hištīg* 'id.' and MP^{Z} ***x****išt* < **hišt* 'id.', already with *x-*), comparable to cases such as OP <u-> 'good' (cf. Av. *hu-*, MP *hu-* 'id.') and <ušk> 'dry; mainland' (cf. Av. *huška-* 'dry', MP *hušk* 'id.')[63].

A more challenging question arises if we accept the proposed hypothesis (i.e. AE *iš*-CV representing OP *is/š*C- or *əs/š*C- < **s/š*C-): why is the assumed prothetic *i-* not consistently reflected in the Pahlavi script as a historical spelling? This remains an open question that can only be addressed through a detailed investigation dedicated to this matter. However, if this interpretation proves to be accurate, it then implies that the development in question traces back to OP. This aligns more closely with the fact that this innovation spread widely beyond Persian.

3.2.4. Other Sources

The following section presents brief observations drawn from additional sources, including Iranian words in Armenian and Syriac, as well as relevant discussions by Islamic linguists from earlier centuries.

While these sources provide valuable insights, their integration into our discussion presents certain challenges. In particular, Iranian words in Armenian and discussions by Islamic linguists pose significant difficulties and cannot be readily incorporated into our arguments without detailed analysis—an endeavor that lies beyond the scope of this paper. A more efficient approach might be to have specialists in the relevant fields examine the information provided by these sources through the lens proposed here,

---

[62] Such a phenomenon is not improbable. We are already aware of some deficiencies (or particular orthographical conventions) of the OP script, wherein certain phonemes were deprived of being written in given conditions. For instance, nasals are not written before certain consonants, cf., e.g. <gdar> *Ga**n**dāra-* in Schmitt's transcription system (see Schmitt 2008: 79-80; 2014: 180). A relevant matter to be noted is that the OP script did not encompass a comprehensive set of signs for all phonemes of the language (cf. Aliyari Babolghani 2024, regarding the dual phonetic value of the OP sign <θ>).

[63] A known orthographical convention to render *hi*C- in the OP script is <hC->, however, this is not fully systematic (see Schmitt 2008: 80).

particularly the idea that the pronunciation of type (a) words with prothesis was universal in MP, rather than confined to MP^{M}.

Armenian: Iranian words of type (a) in Armenian are predominantly recorded with the initial consonant cluster (e.g. Armn. *šnorh* 'grace, gratitude', cf. Prt^{M}, MP^{M} *išnōhr* 'id.'; *spitak* 'white', cf. Prt^{M}, MP^{M} *ispēd* 'id.'), and only occasionally with prothesis (e.g. Armn. *aspar* 'shield', cf. Prt^{M}, MP^{M} *ispar*). The chronology and precise source of these borrowings cannot be determined in many cases. However, it is known that they are primarily borrowed not from Persian but from Parthian and some other non-Persian language(s). For the cases pertinent to our discussion, those with Persian provenance (whether authentic or borrowed) are difficult to distinguish. Furthermore, I am uncertain whether all forms with the initial consonant cluster, regardless of their provenance, reflect the presence of the cluster in the Iranian language from which they were borrowed, or alternative interpretations, such as Armenian adaptation, should be considered, cf. the omission of the original initial vowel in Armn. *štr* (besides *ištr*) 'camel' (cf. Av. *uštra-* 'id.')[64].

Syriac: Similar challenges may be encountered when analyzing Iranian words in Syriac. However, the situation is less complex here, as most of these words are borrowed or quoted from MP[65]. In contrast to Armenian, Iranian words of type (a) in Syriac are predominantly written with prothesis and only rarely with the initial consonant cluster, e.g. Syr. <ʾsph> 'army' (cf. MP^{M} *ispāh* 'id.'); Syr. <ʾsphbyd> and <sphbyd> 'general, commander' (cf. MP^{Z} <spʾhpt'> 'id.'); Syr. <ʾsprmkʾ>, <ʾsprmqʾ>, and <sprmqʾ> 'basil' (MP^{M} *isprahmag* 'flower'); Syr. <ʾspydpqʾ> 'white broth' (cf. MP^{Z} <spyt'pʾk'> 'curd soup', and MP^{M} *ispēd* 'white'); and Syr. <ʾsṭbrgʾ> 'silk dress'[66] (cf. MP^{Z} <stplk'> 'shot silk', and also Arabic *istabraq* 'silk, brocade'[67]).

An especially noteworthy case is Syr. <ʾsṭwnʾ> 'column', which was inherited from and already attested in Official Aramaic, so it was borrowed not from MP *istūn* but from OP <st^{u}una>[68] (cf. § 3.2.3, esp. AE ^{AŠ}*iš-du-na-um*).

Islamic linguists: In their discussions of the initial consonant cluster in Arabic, Islamic linguists have, in some cases, also commented on the same

[64] For the cited Armenian words and discussions relevant to the Iranian loanwords in Armenian, see Schmitt and Bailey 1986.
[65] See Ciancaglini 2008: 11, 14, 37-42.
[66] For the cited Syriac words, see Ciancaglini 2008: 41, 73, 86-87, 110-112.
[67] See Cheung (2016: 3-4, 20-22, 24, 26). He (ibid.) states that Arabic *istabraq* is probably a direct borrowing from EMP *stabrak* 'shot silk' rather than via Syriac.
[68] See Ciancaglini 2008: 30, 70, 110.

issue in Persian[69]. These accounts, however, do not offer a clear or consistent understanding of the issue. Moreover, some of these interpretations appear to be affected by the presumption that initial consonant clusters are universally impossible in any language. Nonetheless, a few discussions that are more pertinent to our subject are as follows—though it should be noted that these discussions are fairly general and not specifically confined to the phonetic context under our consideration.

The author of *Yawāqīt al-ʿUlūm wa-Darārī al-Nujūm* (6th Hijri, 12th century) quotes from a certain Xālidī Naxjawānī, who asserts that contrary to common conception ("mardum pindārand"), the Persian ("Pārsī") words <škm> 'belly' and <štr> 'camel' feature an initial consonant cluster ("awwališān sākin ast"). However, the author strongly disagrees with this statement. He cites Sībūya (2nd Hijri, 8th century), who argued that the initial consonant cluster is beyond human linguistic capabilities. The author further discusses that Xālidī Naxjawānī's misperception stems from the fact that *š* is a fricative(?) ("tanaffusī") consonant, preceded by an implied alif ("alif-ē dar awwal-i ān muqaddar ast"), which occasionally surfaces, resulting in the pronunciations <ʾškm> and <ʾštr>. When the alif is not explicitly manifested, *š* is pronounced after an implied alif ("bar taqdīr-i alif, šīn bigūyad"), leading to the impression of a consonant cluster with *š* ("gumān barand ki šīn sākin gufta ast")[70].

Similarly, Šams-i Qays (6th-7th Hijri, 13th century) asserts that the general consensus among linguists is that initial consonant clusters ("ibtidā ba sākin") are universally impossible in any language. He further notes that Ibn-i Durustūya (3rd-4th Hijri, 9th-10th century) incorrectly held the contrary view, merely based on the observation of certain words pronounced *rubūda* by Iranians ("ʿAjam"), viz. the first consonant in these words is pronounced with an implicit vowel sound between *fatḥa* and *kasra*, as found in *f* in <fɣʾn>, *d* in <drm>, *s* in <srʾy>, and *š* in <šmʾr>—only the latter, meaning 'count', is relevant to our discussion[71].

4. Date of Occurrence

The addition of the prothetic vowel to **s/šC-* (type a), as a general development in several Western Iranian languages, should have commenced in the Early MIr. period (if not earlier, cf. § 3.2.3).

Persian also undergoes a secondary innovation, namely *Vs/šC-* (both types a and b) > *s/šVC-*, which makes it diverge from the other SWIr. The

[69] For a summary, see Ṣâdeqī 1380/2001: 11-13.
[70] *Yawāqīt al-ʿUlūm wa-Darārī al-Nujūm*: 172.
[71] *Al-Muʿjam fī Maʿâyīr-i Ašʿâr al-ʿAjam*: 60-61, fn. 4.

presence of numerous paradigms already spelled without the initial vowel in ENP would tell us that this development, i.e. the forward shift of the prothetic vowel and breaking of the consonant cluster in Persian, might have commenced in the first century after Islam or even before, in the Late MIr. period. Indeed, the form *s**i**pās-dār* 'grateful' (see § 3.2.2), as attested in the Pahlavi text *Andarz ī Pōryōtkēšān*, supports this earlier dating[72].

5. Conclusion

My analysis of the sound change in question can be summarized as follows:

OP *s/šC-* (type a) turns into *is/šC-* in EMP (if not earlier), representing a universal development in MP rather than being restricted to MP^{M}. This development is not unique to Persian; it denotes a broader phonological evolution that likely occurred across various West Iranian languages, probably including all SWIr.

Based on the arguments presented, I suggest that words of type (a) should be transcribed with the prothetic *i-* (e.g. *istārag* as in MP^{M} rather than *stārag*) in Pahlavi (MP^{Z} and MP^{I}) as well. However, in late Pahlavi texts, some words of type (a) may have already been pronounced with an anaptyctic vowel (cf. *s**i**pās-dār* in § 3.2.2). Additionally, the transcription of certain Pahlavi words of type (b) may also require revision (cf. § 3.2.1).

Since MIr. onwards, the sequence *is/šC-* < *s/šC-* (i.e. type (a), e.g. MP *iškamb* 'belly') converges with the other type of initial sequence V*s/šC-* (i.e. type (b), inherited from the earlier period, e.g. MP *uštar* 'camel'), in a similar phonetic context. Thus, from this point onward, they undergo a shared development irrespective of their origin. SWIr. other than Persian generally maintain the structure of this sequence. In contrast, Persian undergoes a secondary change by shifting the prothetic vowel of V*s/šC-* forward, resulting in *s/šVC-*. This development may have begun in Late MIr., continuing into the Early NIr. Accordingly, the presence of 'dual spellings' in ENP (as seen in *iškam ~ šikam*), does not represent dialectal variation; instead, it reflects an ongoing development that ultimately results in NP *s/šVC-* (e.g. *šikam* > *šekam*).

[72] MP^{Z} *zuwān* 'tongue' (in *Ardā Wirāz Nāmag* 57: 1, 63: 3, etc., see Gignoux 1984: 277), the more recent form of *uzwān* (cf. MP^{M} *izwān*; Prt^{M} *izβān*), as well as MP$^{I, M, Z}$ *ruwān* 'soul', the more recent form of MP^{M} *arwān* (cf. Prt^{M} *ruwān* < Prt$^{I, M}$ *arwān* 'id.'; Av. *uruuan-* 'id.'), do not belong here. However, they may indirectly indicate the pre-Islamic age of this type of sound change.

However, monosyllabic words appear to be exceptions to the rule, as seen in examples like NP *ast* 'is' (MP *id.*) and *asp/b* 'horse' (< MP *asp* 'id.'). Moreover, certain words, mostly those starting with the syllables *a/ust-*, have occasionally resisted the development, cf. NP *ostoxân* 'bone' (< MP *astuxān* 'id.'); NP *astar* (sporadically, also ENP *satar*) 'mule' (< MP *astar* 'id.'[73]); NP *ost(o)vâr* (sporadically, also ENP *sotwār*) 'firm' (< MP *awestwār* 'id.'). However, there are also instances of this kind adhering to the rule, such as NP *setordan* 'to erase, shave' (ENP *usturdan* 'id.') and ENP *sitān* 'laying on the back, starfish (sleeping position)' (cf. Av. *ustāna-zasta-*, *ustānāiš... zastāiš* 'with outspread/outstretched hands (in prayer)' translated into MP^{Z} *ustān-dastīh* 'id.'[74]).

The treatment of *s/šC-* in later loanwords, such as those from Western languages, warrants brief mention here. For instance, in the NP of Tehran, such words consistently take a prothetic *e-*, as in English 'standard' > ***e****stândârd* and 'sport' > ***e****sport*. Similarly, in the NP of Kabul, forms like ***e****standard* appear with a prothetic *e-*, although *s****i****port* also occurs. These examples indicate recent and independent developments of initial consonant clusters *s/šC-*. They evidently cannot be conflated with the final phase of the Persian sound change under discussion, specifically V*s/šC-* (both types a and b) > *s/š*VC-, which occurred centuries earlier and in a distinct context.

ABBREVIATIONS

Armn. : Armenian
AB : Achaemenid Babylonian
AE : Achaemenid Elamite
Av. : Avestan (Gathic or Young)
Dez. : Dezfūl
EMP : Early Middle Persian
ENP : Early New Persian
ENP^{J} : Early Judaeo-Persian
ENP^{M} : Manichaean ENP
MIr. : Middle Iranian (period)
MP : Middle Persian
MP^{I} : Inscriptional Middle Persian
MP^{M} : Manichaean Middle Persian
MP^{Z} : Zoroastrian Middle Persian
NB : North Baškardi
NIr. : New Iranian (period)
NL : Northern Lori
(Modern) NP : New Persian
NWIr. : 'Northwestern Iranian'
OIr. : Old Iranian (period)
OP : Old Persian
PF : Elamite Persepolis Fortification
Prs. : Persian in general
Prt^{I} : Inscriptional Parthian
Prt^{M} : Manichaean Parthian
SL : Southern Lori
SWIr. : 'Southwestern Iranian'
Syr. : Syriac
Šūš. : Šūštar

[73] In *Wizīdagīhā ī Zādspram* 3: 58 (see Gignoux and Tafazzoli 1993: 50, 51, 358).
[74] In *Yasna* 29: 5 (see Malandra and Ichaporia 2013: 29, 187, 208).

BIBLIOGRAPHY

Aliyari Babolghani (ʿAlīyârī Bâbolqânī), S. (1396 Š/2017), *Ganǰīne-ye gūyešhâ-ye Īrânī: Haft gūyeš az ḥâšīye-ye Zâgros*. Tehran.

Aliyari Babolghani, S. (2024), "Old Persian <θ> /θ_1, θ_2/: Phonetic Value(s) and Phonological Development(s) into Middle Persian". In: M. Afshin-Vafaei and P. Firoozbakhsh (eds.), *Studia Persica* 23. Tehran.

Aliyari Babolghani, S. (forthcoming) "The Imperfective Marker *at-*, *da-*, *a-*, *ī-*, etc. and its Significance in the West Iranian Historical Dialectology".

Al-Muʿjam fī Maʿâyīr-i Ašʿâr al-ʿAjam by Šams-i Qays-i Rāzī. M. Qazvīnī (ed.), revised by M.T. Modarres Rażavī and S. Šamīsâ, 1388 Š/2009. Tehran.

Amouzgar, J. and A. Tafazzoli (2000), *Le cinquième livre du Dēnkard* [Studia Iranica, Cahier 23]. Paris.

Anonby, E.J. (2015), "The Keshmi (Qeshmi) Dialect of Hormozgan Province, Iran: A First Account". *Studia Iranica* 44/2: 165-206.

Anonby, E.J. and P. Yousefian (2011), *Adaptive Multilinguals: A Survey of Language on Larak Island* [Acta Universitatis Upsaliensis, Studia Iranica Upsaliensia 16]. Uppsala.

Anvarī, Ḥ. ed. (1382 Š/2003), *Farhang-e bozorg-e soxan*, vol. 1, 5. Tehran.

Aṣγarī, S. A-M. (1401 Š/2022), *Gūyeš-e Komzârī-e ǰazīre-ye Lârak: Vâǰšenâsī, dastūr, vâženâme, gozīde-ye motūn*. MA thesis, University of Tehran [unpublished].

Authier, G. (2012), *Grammaire juhuri, ou judéo-tat, langue iranienne des Juifs du Caucase de l'est* [Beiträge zur Iranistik 36]. Wiesbaden.

Barbera, G. (2005), *Lingua e cultura a Minâb (Iran sud-orientale): Profilo grammaticale, testi e vocabolario*. PhD dissertation, University of Naples "L'Orientale" [unpublished].

Bartholomae, Ch. (1904), *Altiranisches Wörterbuch*. Straussburg.

Bailey, H.W. (1979), *Dictionary of Khotan Saka*. London-New York-Melbourne.

Borjian, H. (2016), "Kerman xvi. Languages". *Encyclopædia Iranica* (online edition), available on <KERMAN xvi. LANGUAGES – Encyclopaedia Iranica (iranicaonline.org)> (accessed on May 2022).

Bivar, A.D.H. and M. Boyce (1998), "Eṣṭaḵr". *Encyclopædia Iranica* (online edition, 2012), available on <EṢṬAḴR – Encyclopaedia Iranica (iranicaonline.org)> (accessed on August 2024).

Cheung, J. (2007), *Etymological Dictionary of the Iranian Verb* [Leiden Indo-European Etymological Dictionary Series 2]. Leiden-Boston.

Cheung, J. (2016), "On the (Middle) Iranian borrowings in Qurʾānic (and pre-Islamic) Arabic". halshs-01445860, available on <document (hal.science)>.

Ciancaglini, C.A. (2008), *Iranian Loanwords in Syriac* [Beiträge zur Iranistik 28]. Wiesbaden.

de Blois, F. (2006), "Glossary to the New Persian Texts in Manichaean Script". In: F. de Blois and N. Sims-Williams (eds.), *Dictionary of Manichaean Texts*, vol. II: *Texts from Iraq and Iran (Texts in Syriac, Arabic, Persian and Zoroastrian Middle Persian)*, pp. 89-120. Turnhout.

Durkin-Meisterernst, D. (2004), *Dictionary of Manichaean Middle Persian and Parthian* [Dictionary of Manichaean Texts 3/1]. Turnhout.

Firoozbakhsh, P. (2019), *The Former Dialect of Šīrāz in the Poetry of Šams Son of Nāṣir of Šīrāz (d. 763 AH/ 1362 CE)*. PhD dissertation, the University of Hamburg [unpublished].

Firoozbakhsh (Fīrūzbaxš), P. (2024), "Tarǰome-ye Fârsī-ye Fâteḥat-olketâb mansūb be Salmân-e Fârsī". In: M. Afshin-Vafaei and P. Firoozbakhsh (eds.), *Studia Persica* 23. Tehran.

Gharib, B. (1995, repr. 2004), *Sogdian Dictionary: Sogdian-Persian-English*. Tehran.

Ghilain, A. (1939), *Essai sur la langue parthe. Son système verbal d'après les textes manichéens du Turkestan oriental* [Bibliothèque du Muséon 9]. Leuven.

Gignoux, Ph. (1972), *Glossaire des inscriptions pehlevies et parthes* [Corpus Inscriptionum Iranicarum, Supplementary Series 1]. London.

Gignoux, Ph. (1984), *Le livre d'Ardā Vīrāz: Translittération, transcription et traduction du texte pehlevi*. Paris.

Gignoux, Ph. and A. Tafazzoli (1993), *Anthologie de Zādspram: Édition critique du texte pehlevi traduit et commenté* [Studia Iranica, Cahier 13]. Paris.

Hallock, R. (1969), *Persepolis Fortification Tablets* [Oriental Institute Publications 92]. Chicago.

Henning, W.B. (1933), "Das Verbum des Mittelpersischen der Turfanfragmente". *Zeitschrift für Indologie und Iranistik* 9: 158-253. Republished in 1977, *W. B. Henning Selected Papers 1* [Acta Iranica 14: Hommages et Opera Minora 5], pp. 65-160. Leiden.

Henning, W.B. (1962), "Persian Poetical Manuscripts from the Time of Rūdakī". In: W.B. Henning and E. Yarshater (eds.), *A Locust's Leg: Studies in Honour of S. H. Taqizadeh*, pp. 89-104, pl. iv-v. London. Republished in 1977, *W. B. Henning Selected Papers 2* [Acta Iranica 15: Hommages et Opera Minora 6], pp. 559-574. Leiden.

Henning, W.B. (1965), "A Sogdian God". *Bulletin of the School of Oriental and African Studies* 28/2: 242-254. Republished in 1977, *W. B. Henning Selected Papers 2* [Acta Iranica 15: Hommages et Opera Minora 6], pp. 617-629. Leiden.

Horn, P. (1898-1901, repr. 1974), "Neupersische Schriftsprache". In: W. Geiger and E. Kuhn (eds.), *Grundriss der iranischen Philologie* 1/2: 1-200. Strasburg.

Ḥasandūst, M. (1393 Š/2014), *Farhang-e rīšešenâxtī-e zabân-e Fârsī*, vol. 1, 3. Tehran.

Jaafari-Dehaghi, M. (1998), *Dādestān ī Dēnīg, I: Transcription, Translation and Commentary* [Studia Iranica, Cahier 20]. Paris.

Kellens, J. (1995), *Liste du verbe avestique*. Wiesbaden.

Korn, A. (2005), *Towards a Historical Grammar of Balochi* [Beiträge zur Iranistik 26]. Wiesbaden.

Lazard, G. (1963), *La langue des plus anciens monuments de la prose persane* [Études Linguistiques 2]. Paris.

Lenepveu-Hotz, A. (2011), "Sipāh-i Ispahān: devenir des groupes consonantiques initiaux moyen-perses en persan". *Faits De Langues* 38/1: 79-95.

MacKenzie, D.N. (1990 rev. edit., first 1971), *A Concise Pahlavi Dictionary*. London.

Malandra, W.W. and P. Ichaporia (2013), *The Pahlavi Yasna of the Gāthās and Yasna Haptaŋhāiti*. Wiesbaden.

Mayrhofer, M. (1973), *Onomastica Persepolitana: Das altiranische Namengut der Persepolis-Täfelchen* [Sitzungsberichte der Österreichischen Akademie der Wissenschaften. Phil.-hist. Klasse 286]. Wien.

Morgenstierne, G. (2003), *A New Etymological Vocabulary of Pashto* (compiled and edited by J. Elfenbein, D. N. MacKenzie and N. Sims-Williams) [Beiträge zur Iranistik 23]. Wiesbaden.

Nīknafas Dehqânī, E. (1377 Š/1998), *Barrasī-e gūyeš-e Jīroft va Kahnūǰ*. Kerman.

Paul, L. (2013), *A Grammar of Early Judaeo-Persian*. Wiesbaden.

Pisowicz, A. (1985), *Origins of the New and Middle Persian Phonological Systems*. Cracow.

Qâʾemmaqâmī, A-R. (1401 Š/2022), "Eṣlâḥ-e ʿebârat-ī az *Andarz ī Pōryōtkēšān*". In: *Čehel goftâr dar farhang va târīx va adab-e Îrân*, pp. 401-405. Tehran.

Ravâqī, ʿA. (1381 Š/2002), *Ḏeyl-e farhanghâ-ye Fârsī* (in collaboration with M. Mīršamsī). Tehran.

Rezai Baghbidi, H. (2017), *Middle Persian Historical Phonology*. Osaka.

Salâmī, ʿA-N. (1383 Š/2004), *Ganǰīne-ye gūyeššenâsī-e Fârs 1: Gūyešhâ-ye Davânī, Dahleī, ʿAbdūyī, Kâzerūnī, Kalânī, Kandeī, Kūzargī, Mamasanī, Mâsarmī*. Tehran.

Salâmī, ʿA-N. (1384 Š/2005), *Ganǰīne-ye gūyeššenâsī-e Fârs 2: Gūyešhâ-ye Banâfī, Pâpūnī, Dūsīrânī, Rīčī, Somγânī, Kalânī (Tâǰīkī), Gorganâyī/ Gâvkošakī, Mosqânī, Nūdânī*. Tehran.

Salâmī, ʿA-N. (1385 Š/2006), *Ganǰīne-ye gūyeššenâsī-e Fârs 3: Gūyešhâ-ye Balyânī, Bīrovakânī, Ḥayâtī (Dowlatâbâdī), Dâdenǰânī, LorDârengânī, Dorūnakī/ Mehbūdī, Dežgâhī/ Gowrī, Korošī, Korūnī*. Tehran.

Salâmī, ʿA-N. (1386 Š/2007), *Ganǰīne-ye gūyeššenâsī-e Fârs 4: Gūyešhâ-ye Asīrī, Ahelī, Tangkīšī, Xonǰī, Zâxorūyeī, Šūrâbī, Qalâtī, Kârīyânī, Gerâšī/Zeynalâbâdī*. Tehran.

Salâmī, ʿA-N. (1388 Š/2009), *Ganǰīne-ye gūyeššenâsī-e Fârs 5: Gūyešhâ-ye Aškanânī/Pâqalâtī, Evazī, Banârūyeī, Bīxeī (Beyramī, Bâlâdehī, Kerīškī, Čâhvarzī), Fedâγī, Fīšvarī, Kūrdehī, Galledârī, Lârī*. Tehran.

Schmitt, R. (2008), "Old Persian". In R. Woodard (ed.), *The Ancient Languages of Asia and the Americas*, pp. 76-100. Cambridge.

Schmitt, R. (2014), *Wörterbuch der altpersischen Königsinschriften*. Wiesbaden.

Schmitt, R. and H.W. Bailey (1986), "Armenia and Iran iv. Iranian influences in Armenian Language". *Encyclopædia Iranica* (online edition, 2011), available on <ARMENIA AND IRAN iv. Iranian influences – Encyclopaedia Iranica (iranicaonline.org)> (accessed on May 2022).

Suleymanov, M. (2020), *A Grammar of Şirvan Tat* [Beiträge zur Iranistik 46]. Wiesbaden.

Sundermann, W. (2003), "Ein manichäischer Lehrtext in neupersischer Sprache". In: L. Paul (ed.), *Persian Origins: Early Judaeo-Persian and the Emergence of New Persian*, pp. 243-274. Wiesbaden.

Ṣâdeqī, A-A. (1357 Š/1978), *Takvīn-e zabân-e Fârsī*. Tehran.

Ṣâdeqī, A-A. (1380 Š/2001), "Taḥavvol-e xūše-ye ṣâmet-e âγâzī". In *Masâʾel-e târīxī-ye zabân-e Fârsī*, pp. 11-23. Tehran.

Ṭâherī, E. (1389 Š/2010), *Gūyeš-e Baxtīârī-e Kūhrang*. Tehran.

Ṭâherī, E. (1395 Š/2016), *Gūyeš-e Lorī-e Boyeraḥmadī*. Tehran.

Utas, B. (1968), "The Jewish-Persian Fragment from Dandān-Uiliq". *Orientalia Suecana* 17: 123-136.

Walker, J. (1956), *A Catalogue of the Arab-Byzantine and Post-Reform Umaiyad Coins* [A Catalogue of the Muḥammadan Coins in the British Museum 2]. London: The Trustees of the British Museum.

Xâleqī-Moṭlaq, J̌. (1398 Š/2019), *Vâǰšenâsī-ye Šâhnâme: Pažūhešī dar x^{w}âneš-e vâžegân-e Šâhnâme*. Tehran.

Yawāqīt al-ʿUlūm wa-Darārī al-Nujūm. M.T. Dânešpažūh (ed.), (1345 Š/1966), Tehran.

Zadeh, T. (2015), "The *Fātiḥa* of Salmān al-Fārisī and the Modern Controversy over Translating the Qur'ān". In: S.R. Burge (ed.), *The Meaning of the Word: Lexicology and Qur'anic Exegesis*, pp. 375-420. Oxford.

Zhang Zhan (2023), "Two Judaeo-Persian Letters from Eighth-Century Khotan". *Bulletin of the Asia Institute* 31: 105-133.

Zhang Zhan and Shi Guang (2008), "A Newly-discovered Judaeo-Persian Letter". *Journal of the Dunhuang and Turfan Studies* 11: 71-99 (in Chinese).

JIL 1 (2024): 90-117

The Khonji Dialect of Lārestān

Habib Borjian
Rutgers,
The State University of New Jersey

doi.org/10.46991/jil/2024.01.05

Abstract: Southwest Iranian languages are significantly underrepresented in academic literature, lacking comprehensive descriptions. This paper aims to fill this gap by examining the Lārestāni language variety spoken in Khonj, which is otherwise poorly known. In phonology, it is shown that uvular stops and fricatives have only recently infiltrated the Khonji sound system. Khonji stands out in the construction of noun phrases due to the presence of a deictic suffix and an intricate system of adpositions. The verb inflection in Khonji is notable for its ten identified aspectual and modal affixes. Combinations of these affixes give rise to a system of five simple tenses (surpassing the previously identified four), alongside corresponding progressive tenses, supplemented by at least four identified subjunctive tenses. The binary stative-dynamic distinctions in posture verbs, as illustrated in a diagram for 'sit', align more closely with English than Persian. Ergativity is prominently manifested through two sets of person markers, creating phrases similar to Middle Persian and counterintuitive to New Persian. Experiencer constructions characterize the expressions of possession and modals in the language.[1]

Keywords: Southwest Iranian, phonology, morphosyntax, subjunctive, ergative, stative verbs

Habib Borjian
E-mail: hb146@columbia.edu

Received: 15.03.2024
Revised: 10.07.2024
Accepted: 20.07.2024

This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License

Introduction

[1] I am deeply grateful to the anonymous peer reviewers of this paper for their useful comments.

1. Introduction

Khonj is a township with approximately 20,000 residents, located 110 km north of the Persian Gulf and 270 km south of Shiraz, the capital city of Fārs province. Situated at the northwestern corner of Lārestān, Khonj lies in a subtropical region, traditionally classified in Persian climatical zoning as *garmsirāt*, where dates and citrus can be farmed. In early modern history Lārestān[2] formed a prosperous and industrious province extending south of Fārs down to the Persian Gulf. See the map in Figure 1, which shows close agreement with the isogloss map in Borjian 2020. Today, Lārestān is divided between two provinces and fragmented into several distinct sub-provinces,[3] with Khonj being one of them.

Khonji, known to its speakers as *xiňǰi* or *xoňǰi*, is spoken in Khonj and its surrounding villages. It belongs to the Lārestāni Language Group (Molčanova 1977), which, together with the Garmsiri Language Group of historical Kermān, forms the larger "Garmsiri" family (Borjian 2017). Lārestāni is known to outsiders as *ačomi*, derived from the word *ačom* 'I go,' which is characteristic of the Lārestāni Language Group. Lārestāni is spoken in dozens of settlements, including Lār, the historical seat of Lārestān.

A significant amount of data is published on individual or groups of Lārestāni dialects, mostly by local pundits. However, no detailed, rigorous study exists on the dialectal subdivisions of the group. My focus here is not on Lārestāni in general, which requires a thorough typological study, but specifically on the Khonji dialect. This focus aims to understand the structure of a single variety before moving forward to a comparative study of the language as a whole.

The primary source of Khonji data is the two editions of the monograph published by Loṭf-ʿAli Xonji (1999, 2009).[4] The former edition is particularly valuable for its texts, despite being translated from Persian works. It received a scholarly review by Ṣādeqi (2003), and its data was utilized by Dabir-Moqaddam

[2] For history, see Calmard 1986.

[3] I tend to overlook new geographical divisions, as historical toponyms best describe the geography where Iranic languages are formed.

[4] In 2015, I conducted interviews with Loṭf-ʿAli Xonji (Khonji) regarding his documentation and obtained additional data through telephone conversations. The sentences without citation are those I elicited from him. Mr. Khonji had a distinguished career as a senior anchorman at BBC Persian. Fluent in English and French, in addition to his native Persian, he often reflected on the distinctive nature of his mother tongue, Khonji, compared to these languages. Although not a linguist by profession, he developed a deep understanding of linguistics while compiling his books on Khonji.

(2014: §§5.3.10-13) in his two-volume work on Iranian typology, which focuses on the tense-based split alignment in the syntax.

Kamioka et al. (1986) published a pioneering Khonji (together with Lāri) glossary, of 1,000 items, accompanied by a phonology. In a series on the Fārs province by Salāmi,[5] volume IV (2007) includes words and elicited sentences of Khonji together with eight other dialects of the province; his Khonji data differs from the abovementioned documentations in significant phonological (e.g., q > k) and morphological (e.g., 2sg. verbal ending *-e* for *-eš*) traits, indicating substantial micro-dialectal variation within Khonji due to areal, generational, and social distribution. Khonji linguistic materials are also found in Eqtedāri (1955), Vosuqi (1995: 98-99, 173-178), Kalbāsi (2009: 265-267), and A.-Ḥ. Xonji (2015).

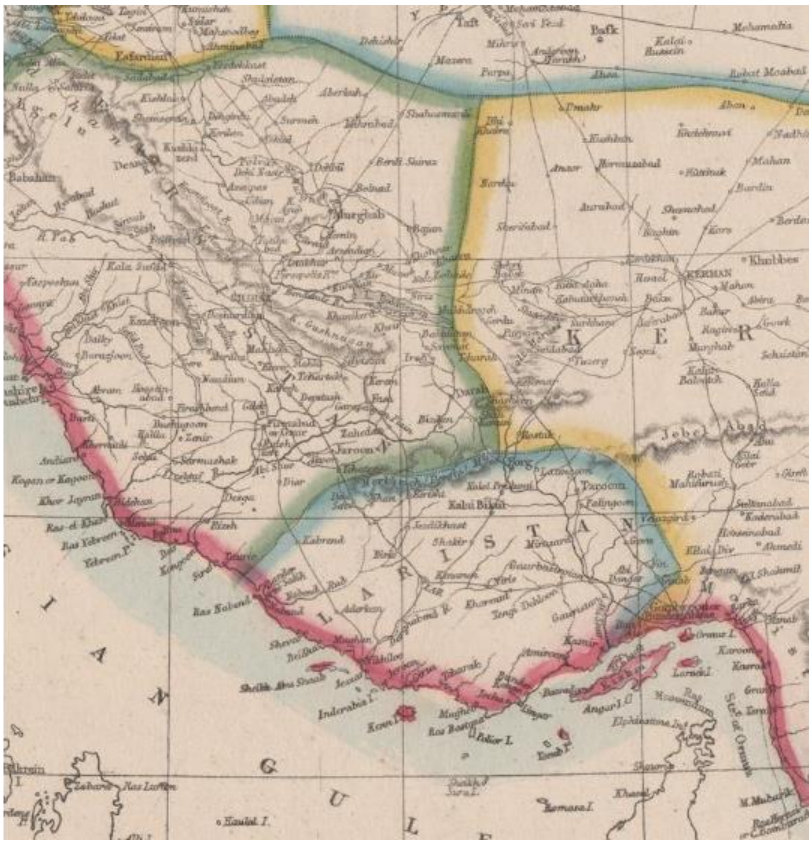


Fig. 1. Lārestān province in the 19th century, located between Fārs and Kermān provinces and the Persian Gulf. Source: "Persia", by Edward Weller, 1863 (author's collection).

[5] For a discussion about Salāmi's elicited data, see §24, below.

2. Consonants

The genuine inventory is probably /p b t d č ǰ k g f v s z š x h m n r l y/. Most notable differences with modern Persian occur in back consonants:

/h/ has lost phonemic status at onset: *(h)ow* 'water', *(h)awr* 'cloud', *(h)esm* 'name', *(h)omā* 'we', *ā(h)en* 'iron'.

/k, g/ have no palatalized allophones as in Persian.

The voiced uvular fricative *ḡ* occurs but rarely in the data, as in *čāḡ* 'fat' and *lāḡar* 'lean', implying influence by Persian pronunciation. Otherwise, Pers. *ḡ* (Arabic and Turkic loans included) is regularly realized as /x/: *morx* 'hen', *birix* 'ewer', *kalāx* 'crow', *xeč* 'ram', *xura* 'unripe grape', *šalxom* 'turnip', *xarbal* 'sieve', *kāxoz* 'paper', *demāx* 'nose', *portexāl* 'orange', *čaxe* 'knife', *čomāx* 'club', (Arabic) *xam* 'sorrow', *xossa* 'grief' (also *ḡossa*), *xark* (< ḡarq) 'drowned'.

The voiceless uvular stop *q* seems recent in Khonji; it occurs in free variation with /k/ in data from Kamioka et al. and L. Xonji, but realized as *k* in a big majority of words in Salāmi's data: *kad* 'length', *čekad* 'how many', *kand* 'sugar cube', *kermez* 'red', *akik* 'agate', *nokra* 'silver', *bark* 'lightening', *vakti* 'when', *mowke* 'time', *fakat* 'only', *āšek* 'in love', *kalom* 'pen', *kolf* 'padlock', *kollāb* 'hook', *keyči* 'scissors', *monkāš* 'tweezers', *boškāb* 'plate', *kašox* 'spoon', *kabloma* 'pot', *sakf* 'ceiling', *kannāt* (Pers. qanāt) 'subterranean aqueduct', *ka:va* 'coffee', *hokuk* 'wages', *hakikat* 'truth'. The recent currency of some of these words in colloquial Persian suggests a synchronic status of this sound shift in Khonji; meaning that at least some speakers perceive uvular plosive as velar.

The interdental fricative *δ* occurs postvocalically in Salāmi's data, e.g. *āδam* 'person'; its inconsistent usage suggests that some of his eight informants spoke in a hybrid dialect.

A peculiar variation occurs in the segment /st/ in Xonji's data and /ss/ in Salāmi's data. Mr. Xonji explained to me that his version belongs to the polite variation of the dialect. Indeed, a high register is quite plausible even in informal, non-written languages, as I have observed in the Central Plateau languages.

3. Vowels

The substantial variation in the data makes it difficult to bring the vocalic inventory of Khonji in the abstract. Kamioka (1986) defines Khonji vowels as /ī e~i a ā o~u ū/. However, the correspondence between *ī* and *ū* and their short forms seems governed by the tense-lax system known in modern Persian. As such, Khonji vowel phonemes may be defined as simple as /i e a ā o u/ plus diphthongs /ey ay ow aw/, with the following notes:

/ā/ [ɑˑ, ɒˑ] is conditionally in free variation with /a/ [a], as in *vā ~ va* 'by, with', especially when preceding a stressed syllable, e.g., *kāré ~ karé* 'the work'.

/e/ and /i/ are in free variation in some words, resulting in doublets such as *zemi ~ zimi* 'earth', *íngo ~ éngo* 'here', *vilāyat ~ velāyat* 'village'.

/o/ and /u/ are in free variation in some words, e.g., *ko ~ ku* 'where?', *ǰonga ~ ǰunga* 'male', *tof ~ tuf* 'spit', *ǰuhu ~ ǰohu* 'pretty', *dudu ~ dodu* 'tooth'.

Kamioka also defines /ee aa oo/ as vocalic subsets, e.g., prepositon *teʾe ~ tee ~ te: ~ tey ~ te* (*te + ezafe* marker *-e*) 'in'; *deer/de:r* 'door', *vā-deet-/de:t-* (< dōxt-) 'sew'; *ǰomaa* (Salāmi *ǰomaha*) 'Friday'; *xooge* (Salāmi *xoge*) 'sister'.

Vowel elision often occurs in quick speech, when words and morphemes are uttered in a single breath: *čaš-oš_a_nu-fta_š=got ~ čaš-oš a nu ofta, oš=got* '[as] his eye caught the bread, he said…'. (Xonji 2009: 304).

4. Nominal Inflection

Nominal inflectional morphemes include plural *-iā́, yā́,* indefinite *-i, -e*, definite *-é*, deictic *-o*, diminutive *-aká,* ezafe *-e, -y*. Examples:

> *sib* 'apple', *sibiā́* 'apples', *síbi* 'a (certain) apple', *sibé* 'the apple', *me síbo* 'that apple', *sibaká* 'little apple', *síbe sorx* 'red apple'; (in noun phrases) *{sib-iā́}-i* '(certain) apples', *me {sib-iā}-o* 'those apples', *me {sib-e sorx}-o* 'that red apple'
>
> *xuná* 'house', *xunayā́* 'houses', *xunáe* 'a house', *xunaé* 'the house', *xuná-e/-y (h)omā* 'our house', *xunayā́-e gap* 'big houses', *xuná-e* 'it is a house'

Nouns ending in *-i* as a rule inflect and receive the hiatus-breaker *-n-*, e.g.,

maí/mayí 'fish', *mae-n-iā́* 'fish$_{PL}$', *maé-n-i* 'a fish', *mae-n-é* 'the fish', *en maé-n-o* 'this fish', *máy-n-e sorx* 'red fish'.[6]

5. Pronouns and Deixis

There is a single set of freestanding personal pronouns (Table 1), serving as both subject and object, e.g., *amā ondem* 'we came', *amā abene* 'he'll see us'. There is no distinction of gender.

Demonstrative pronouns and adjectives are *e, en* 'this', *me* 'that, this', *on* 'that', *iā, enyā* 'these', *miā, onyā* 'those' (also 'these'); intensives are *hamin/hamon* 'this/that very (same) one'. Demonstrative adjectives co-occur with the deictic suffix *-o*, as in *me ketāb-o ase to_m=xeli* 'I bought that book for you'.

Circumstantial adverbs include *ékā, íngo/éngo* 'here', *ónkā, óngo, mékā* 'there'; *éndā, índo* 'this way', *mándā, ondo* 'that way' (corresponding to Pers. čonin, čonān).

Table 1. Personal pronouns and verbal endings

	Pronouns		Verb Person Markers
	Freestanding	Pron. Clitics	Endings
sg. 1	*mo*	*om*	*-om*
2	*to*	*ot*	*-eš*
3	*on, u*	*oš*	*-e*
pl. 1	*amā, (h)omā*	*mo(n)*	*-em*
2	*šomā*	*to(n)*	*-i*
3	*onyā, ešu*	*šo(n)*	*-et*

6. Pronominal Clitics (PC)

These pronouns (Table 1) are either suffixed or prefixed, showing mobility within the phrase and proclivity to fuse with verb morphemes and prepositions. For instance, the third person singular clitic is realized as *-š, š-, -oš, šo-, oš-*, and *ša-* with prepositions and the verb durative marker; the third person plural is *šo/šu* is *šon-* prevocalically and *šā-* with the durative marker (§16). Phonetic variation in other clitics occurs as well. Pronominal clitics have a wide range of oblique functions:

[6] A contraction of *maé-n-e sorx*.

(1) Possessive (POSS): *mai-t* 'your$_{SG}$ fish', *lu-š* 'his/her face'.

(2) With prepositions (§8): *ša-tek* 'in it', *sa-lu* 'on it', *ša-zel* 'under it', *ša-z* 'from it'.

(3) Direct object hosted by the verb (§16): *m=a-ben-eš* 'you$_{SG}$ (will) see me', *om=nāben-eš* 'you$_{SG}$ do/will not see me'.

(4) Indirect object: *nu oš=hā-t-i* 'give$_{PL}$ him bread'.

(5) Subject (agent) in ergative (AC) (§16): *ot=ded-om* 'you$_{SG}$ saw me'.

(6) Experiencer (XPER), for which see "Possession" (§23) and "Modal Forms" (§25).

(7) Reflexive (REFL) with the base *xo-*: sg. 1 *xom,* 2 *xot,* 3 *xoš,* pl. 1 *xómu,* 2 *xotu,* 3 *xošu.* Functions are reflexive (ex. 1, 11) and emphatic (ex. 2, 6).

(1) *e* *xuna-e* *xarāb-o* *a* *xo-tu* *be-freš-i*
this house-EZ ruin-DEIC PREP REFL-PC.2PL SBJV-sell.PR-2PL
'Sell this ruined house to yourselves.' (Xonji 2009: 63)

(2) *xo-šu* *šo=got* *ke* *xalāf* *šo=kerd-e*
REFL-PC.3PL AC.3PL=say.PST SUB mistake AC.3PL=do.PST-PP
'They themselves said that they have made a mistake.' (Xonji 2009: 63)

7. Object Marking

In the absence of an accusative marker, such as Persian -rā, various strategies are used to mark direct object: (1) SOV word order (ex. 3); (2) Verb agreement with the object in past transitive tenses (*om=kerd-et* in ex. 4); (3) Preposing the particle *ase* (otherwise a preposition) (ex. 5, 6); (4) The stress marking generic object nouns may shift to the verbal ending with specific objects: *qāli afrešé* 'he sells the rug' versus *qālí afreše*[7] 'he sells rugs'; *ketāb om=xelí* 'I bought the book' versus *ketáb-om=xeli* 'I bought books'.[8]

(3) *Hasan* *Ali* *tey* *bāḡ* *oš=di*
PN PN PREP garden AC.1SG=see.PST.CRUSH
'H. saw A. in the garden.' (Salāmi 2007: 324)

[7] The secondary accent on *afreše* was not perceivable.
[8] Xonji 2009: 32.

(4) *yak* *mablağ-i_am* *a farrāš-iā=m*
one amount-INDF_also to servant-PL=AC.1SG
da *o moraxas* *om=kerd-et*
give.PST and discharge AC.1SG=do.PST-3PL[9]
'I also gave some money to the servants and discharged them.' (Xonji 2009: 320)

(5) *ase* *pos-iā* *om=di*
PREP boy-PL AC.1SG=see.PST.CRUSH
'I saw the boys.'

(6) *ase* *xo-š_oš=di*
PREP REFL-PC.3SG_AC.3SG=see.PST.CRUSH
'He saw himself.' (Xonji 2009: 30, 63)

8. Adpositions

Khonji is prepositional, in the sense that an adposition precedes the noun it governs, usually with an *ezafe* connector. However, most prepositions have postclitic forms that are suffixed to pronominal clitics (Table 1). Frequent adpositions are: *a* 'to, from', *az* 'from', *ase,*[10] *-su* 'for', *bā, -(o)mra* 'with', *barā́(e)* 'over', *lā-va* 'together with', *le, lu* 'on', *pás(e)* 'behind', *péš(e), -peynā* 'before, with, next to', *tā* 'till', *táht(e)* 'beside', *tey, té(e)*, *ték(e)* 'in, inside', *zél(e)* 'under'. Examples:

(7) *xuna ase on omxeli ~ xuna ša-su omxeli* 'I bought the house for him'
(8) *taht-e mo honi ~ ma-taht honi* 'sit next to me'
(9) *bā onyā očo ~ šon-omra očo* 'go with them'
(10) *ketāb-ot ša-peynā hod* 'your book was with him'
(11) a_2 *xom* om_1*=got ~* m_1*=*a_2 *xom got* 'I_1 said to$_2$ myself'
(12) *onyā te menserā nehet* 'they are not in the courtyard'
(13) *kolāh le ser-aš nehod* 'he had no hat on his head'
(14) *malaxiā gonomiā lā-va riša šoxo* 'the locusts ate the wheat altogether with roots'

[9] See §9, Ergative.
[10] The preposition *ase* also functions as a particle in marking the direct object (see Object Marking, §7) and in forming a secondary present-future (§24).

9. Particles

Adverbs are formed with *va-*, as *va-xaši* 'happily', *Hasan va-hila goruxt* 'H. fled deceptively'; *-(e)ndā*, as *éndā* 'this way', *mándā,* 'that way', *hámendā* 'as such', *čéndā* 'how'; *e-* (< em-) in temporal *eroz* 'today', *ešow* 'tonight', *esāl* 'this year'; *-in* in *pišin* 'forenoon, noon', *pasin* 'midafternoon'.

Question words include *če* 'what', *ču, čéndā* 'how', *čerā, ase-če, sey-če* 'why', *četay* 'which', *čod* 'how much', *ke* 'who', *kodom* 'which', *ko, ku* 'where', *kay* 'when'.

Note also *hanu* 'yet', *hiǰ* 'none', *gāhi* 'never', *dega, -eydu* 'else, other' (*pos-e dega ~ pós-eydu* 'other boy'), *hā* 'yes', *no* 'no', *_o* 'and'. The comparative marker is *-ta* (seldom *-tar*), e.g., *keyeta* 'smaller', *ǰohuta* 'prettier', *gap-ta_ye* 'it is larger', *gaptar az mo* 'older than me'.[11]

10. Verb Phrase

The dialect has a five-fold system of tenses in the indicative, supplemented by the subjunctive mood. The dual present–past stem pattern is defied by the present progressive, which is built on the past stem. Among personal endings (Table 1), the third singular conjugates aberrantly (Table 2). Transitivity is governed in past tenses by ergativity (Tables 3 and 4). Posture verbs such as 'sit' have punctual and stative aspects, as in English (Table 5).

The complexity of the verb forms led me to decompose them into formative elements (§14) and try to identify the underlying morphemes common to West Iranian. The resulting picture reveals that original compounds and agglutinations have fused into single-word forms, leaving no "periphrastic" verbs in Khonji, which attests to a long process of tense formation in the language. However, new analytical forms are emerging under Persian influence (§24).

11. Stems

The present stem is employed in the present-future, present subjunctive, and imperative. All other tenses, including the present progressive, employ the past stem.

[11] One of the reviewers brought up this point: /r/ resurfaces before vowels, but it is obstructed by the hiatus filler *-y-* in *gapta-y-e*. This matter needs more attention.

The diachronic opposition between inherited present and past stems is markedly diminished. Regularized or secondary past stems, with markers *-ad-* and *-ed-*, such as (present : past) *kar-* : *kared-* 'plant', *feress-* : *feressad-* 'send', *soxen-* : *soxenad-* (causative) 'burn' appear to be in the minority. Stem pairs inherited from Old Iranian, are "irregular" in the sense of showing no synchronic derivational interrelationship, e.g., *gard-* : *gešt-* 'turn', *gel-* : *gelet-* 'get', *ni-* : *šass-* (punctual), *(h)od-* (stative) 'sit'; as well as diachronically suppletive stems *-i-/-r-* : *ond-* 'come', *ben-* : *ded-* (also secondary *ben-ed-*) 'see'. Moreover, a large subset of present stems are historically derived from old past stems, e.g., *xās-* : *xāsed-* 'bite', *xat-* : *xated-* 'sleep', *pox-* : *poxt-* 'cook', *sox-* : *sot-* 'burn' (< *sōxt-), *bi-* : *best-* 'throw'. Partially assimilated present stems include *bon-/bass-* : *bass-*, as in *vā-bon* 'close!', *vā-bass-e* 'he'll close'. Some past stems are truncated when word-final: *ded-/-di* 'see', *kerd-/-ke* 'do' (designated SHORT in interlinear glosses).

Verbal nouns are also employed in verb forms (§14, § 25). The infinitive is the past stem + *-a /-o*, e.g., *šasta*,[12] *šasso*[13] 'to sit'. The past participle marker is *-e(st)/-ess*, with the allomorph *-ez-*.

12. Passive and Causative

Passive stems are marked by *-eh-* : *-eh-est-*, as in *šiša eškahest-Ø* 'the glass broke' ~ *šiša š=eška* 'he broke the glass'. Causative stems are marked by *-en-*: *-en-ad-*, as in intransitive *sox-* : *sot-*, causative *soxen-* : *soxenad-* 'burn', e.g., *xaǰa sox-ā-e* 'the firewood is burning' ~ *xaǰa soxenad-ā-m*[14] 'I am burning firewood'.

13. Preverbs

The only active lexical prefix in Khonji is *vā-*. Its semantic effects are limited to a few verbs, including *xord-* 'eat' ~ *vā-xord-* 'drink'; *košt-* 'kill' ~ *vā-košt-* 'extinguish, turn off'. There are stems that occur only with the preverb: *vā-mon-* : *-mod-* 'stay; lack behind', *vā-bon-* : *-bass-* 'close', *vā-kōven-* : *-kōved-* 'search',[15] *vā-ǰor-* : *-ǰoss-* 'search, find', *vā-doz-* : *-det-* 'sew', *vā-paraven-* 'strain'

[12] Xonji 2009.

[13] Kamioka et al. 1986; Salāmi 2007.

[14] Note the present progressive with the past stem (§17.1).

[15] *kōv°*, an odd outcome of *kāv-, may be analyzed this way: kāv- > (the stem) kō-, suffixed by the filler *-v-*.

(Pers. pāludan). This preverb is also used with light verbs: trans. *gerā vākerdo*, intr. *gerā vābodo* 'to blaze'.

The *vā-*prefix remains attached to the stem in all forms, e.g., *a-vā-xor-eš* 'you$_{SG}$ drink', including negation (*ne-vā-xor-et* 'that they do not drink'), and suppresses the modal prefix *be-* (§14.1), as in *vā-xor-eš* 'that you$_{SG}$ drink', *vāxo* 'drink$_{SG}$!'.

Moreover, *(h)o-* and *(h)ā-* supress modal *be-* only in the subjunctive present and imperative of a subset of verbs: (examples in the subjunctive 1sg.) *ó-čom* 'go' (neg. *ne-čom*), *(h)o-xatom* (Salāmi *ve-xatom*) 'sleep', (*h)o-nesom* 'put', (*h)o-niom* (Salāmi *vi-neyom*) 'sit', *orostom* 'get up', *(h)ā-tom* 'give'. Likewise, the verb *vaystada/vaessado* 'to stand up, to stop', with an original preverb *vā- (cf. colloquial Pers. vāysādan), has the forms *vaysom* 'that I stand', *mavaysi* 'do not stand!'.

14. Aspectual and Modal Affixes

In addition to the stem, preverbs, and person markers (verbal endings and pronominal clitics), the following elements are discernable in verb forms. These can be summarized as subjunctive *be-*, durative *a(d)-*, participal *-est-*, copula stems *b-* and *bost-/bod-*, and the enigmatic *-ā.* These components collectively contribute to the complexity and variation of verb forms in the language.

(1) *be-* marks the subjunctive present and the imperative.

(2) *a-*, *ad-/at-* (before vowels), equivalent to Persian durative marker mi-, marks the present-future, the imperfect, and the progressive forms of perfect tenses. This durative marker prefixes normally to the stem (*a-čed-om* 'I was going', *ad-ār-om* 'I bring'), but may influence the stem, as in *a-ftad-et* 'they would fall' (cf. *oftad-et* 'they fell'), *ad-ānd-om* (*at-ond-om* in Salāmi) 'I was coming' (cf. *ond-om* 'I came'). The marker coalesces into *-ā-* with the negative marker (§20) and with plural pronominal clitics (Table 1), e.g., *mādi* (← mo + a + di) 'we would see' (§16).

(3) *-est-/-ess-,-e* (in final position) marks the perfect and pluperfect. It is thus the past participle formant (< ast 'is') in the context of West Iranian morphology.

(4) *-ez-* is suffixed to the past stem in the past-pluperfect and the subjunctive perfect and pluperfect. It is analyzed (Ṣādeqi 2003: 129) as a reduced form of the past participle *-est-* when the latter coalesces with succeeding /b/, the

stem-initial of the subjunctive and perfect of 'be' (Table 6); hence, *-est-b- > -*ez-b-*.

(5) *-ā-* is suffixed to the past stem in the present progressive. Ṣādeqi (ibid) analyzes this morpheme as the fusion of the infinitive marker *-a* (also defined as *-o*) and the preposition *a* *'in' (otherwise ablative in Khonji, §8). As such, *dedā(o)m* 'I am seeing'[16] would have the underlying form *deda-a-om 'I am in (the process of) seeing'.[17] The existence of a preposition in this position seems rather odd to me. It is more plausible to assume that the inserted -a- is the durative marker (see (2), above), which has oddly moved forward in the morpheme arrangement. Nevertheless, quite tenable is an underlying locative formation based on the infinitive, a structure also found in some of the Garmsiri dialects of Kermān (Borjian 2017: 311), e.g., Minābi *a-kerden-om* 'I am doing' (Barbera 2005).

(6) *-āst-* and *-āz-*. These segments appear in the past progressive, e.g., *čed-āst-od-om* 'I was going'; and in the subjunctive progressive, e.g., *xond-āz-bom* 'I may be reading', *xond-āz-bāš* 'keep reading!'. Ostensibly related to the aforesaid participle *-est-/-ez-*, their role in these imperfective/subjunctive tenses is all but counterintuitive. Ṣādeqi (2003: 132) conjectures a morphological degeneration due to a phonological fusion between the perfective maker and past copulas.

(7) *-bost-,* employed in the past-pluperfect, is the past participle of 'be' (Table 6), functioning here as an original auxiliary synthesized into the verb phrase. It occurs in transitive verbs in its third person singular form *bode* for all persons, but it emerges in full with postclitics (Table 4).

(8) *-boz-* is infixed in the subjunctive pluperfect, e.g. *ond-ez-boz-bom* (Pers. āmada buda bāšam). The underlying morpheme is ostensibly a contraction of *bost-* 'been', thereby the synthesized auxiliary *boz_bom* (Pers. buda bāšam).

(9) *-od-*, *-u* (in final position), employed in the pluperfect, is basically the past stem of 'be', which functions here as auxiliary.

(10) *-b-*, employed by subjunctive complex forms, is the subjunctive stem of 'be', which functions here as auxiliary in a diachronic sense. As such, it conjugates

[16] Salāmi (2007: 250) gives the paradigms with a hiatus filler: *dedā-y-om, dedā-y-e, dedā-y-i* '[dāram] mibinam, etc.'

[17] Cf. Lāri *a-xetā-i* 'he is sleeping', which Molčanova (1982: 433 f.), invoking parallel Tatic forms, parses as the preposition *a-* prefixed to the infinitive; she gives no concrete justification for the existence of *-ā-*.

in intransitive forms and appears invariably as the third singular *be* with transitive stems.

15. Person Markers

The verb personal endings listed in Table 1 merit the following notes. The second person singular ending *-eš*, characteristic to Lārestāni, occurs as *-e* in Salāmi's data. The second person plural ending *-i* becomes *-ay* after *i*-final stems. The third singular is unmarked in past tenses; in the present, it is regularly suffixed with *-e*, as in *anese* 'puts', *ahere* 'lets', *adāre* 'brings'; zero after *i*-final stems: *abi* 'throws', *ani* 'sits'.

Third singular forms are contracted, sometimes beyond recognition, in a subset of Khonji verbs. This behavior, common in other Southwest Iranian languages, is designated as a 'crush' by Ilya Gershevitch (1970), and I will use this term for Khonji.[18] Examples are (1sg ~ 3sg) present-future *a-zen-om ~ a-zot* 'hits', *akenom ~ akot* 'does', *adonom ~ adu* 'gives', *abarom ~ aba* 'carries', *axarom ~ axa* 'eats', *adiom ~ adā* 'comes', *atom ~ ada* 'gives'; past *ondom ~ oma* 'came', *čedom ~ ču* 'went', *xatedom ~ xat* 'slept'. See Table 2 for full paradigms.

Table 2. Conjugations of *onda* 'to come'

	Pres.-Future	Preterit	Perfect	Subjunctive
sg. 1	*adiom*	*ondom*	*ondestom*	*berom*
2	*adieš*	*ondeš*	*ondesteš*	*bereš*
3	*adā*	*oma*	*onde*	*biā*
pl. 1	*adiem*	*ondem*	*ondestem*	*berem*
2	*adiay*	*ondi*	*ondesti*	*beri/biāy*
3	*adiet*	*ondet*	*ondestet*	*beret*

16. Ergativity

Khonji inherits from Middle Persian a tense-based split alignment, that is, accusative in the present and ergative in the past. In the present tense, personal endings agree with the subject. In past tenses, personal endings

[18] For diachronic justification, cf. Gershevitch 1970. For application on the Fārs language group, see Borjian, forthcoming: §5.5.

agree with the patient/object, while the (oblique) pronominal clitics (PC; Table 1) mark the agent/subject.[19]

(15) pres. *šā-ben-em* 'we see them'
past *mo=ded-et* 'we saw them'

In the following examples note false friends with Persian, e.g. 'I greeted him', etc.

(16) *ke* *ǰār=oš* *zat-om*[20]
who call=AC.3SG hit.PST-1SG
'Who did call me? (Salāmi 2007: 329)

(17) *har* *ke* *mo=š* *di,* *salām=oš* *kerd-om*
every person I=AC.3SG see.PST.CRUSH hello=AC.3SG do.PST-1SG
'Whoever saw me, greeted me.' (Salāmi 2007: 333)

(18) *če=tu* *got-om?*
what=AC.2PL say.PST-1SG
'What did you$_{PL}$ tell me?[21]

Since the direct-oblique case system of earlier Middle Persian is lost in Khonji, the agent clitic (denoting obliqueness) is obligatory even with an overt lexical agent: *on amā oš=di* 'he saw us'. The patient marker (verb ending) is optional when the patient is specified: *amā oš=di ~ amā_š=di ~ oš=ded-em* 'he saw us'.

The clitics appear in the following basic forms (for the verb 'see'):

Preterit: AC=see.PST.CRUSH
sg. *om=di, ot=di oš=di*
pl. *mo=di, to=di, šo=di*

Imperfect: AC.DUR=see.PST.CRUSH
sg. *ma=di, ta=di, ša=di*
pl. *mā=di, tā=di, šā=di*

The agent clitic always comes ahead of the stem; it may move off the verb and attach to the direct object, an indirect object, and prepositional and adverbial

[19] For a more detailed study of syntactic alignment in Khonji, see Dabir-Moqaddam (2014): §5.310-313.

[20] Note that *-om* is a verb ending here; it resembles the pronominal clitic of the first person singular.

[21] Note that with the verb 'say', 'me' is treated as *patient* and not an indirect object accompanied by an adposition. This occurs also in other Iranian languages. See Also Xonji 2009: 272 ff.

phrases. Examples:

(19) *sag* *{pā-e* *pos-iā}=š* *xast*
dog foot-EZ boy.PL=AC.3SG wound.PST
'The dog bit$_{PST}$ the boys' feet.' (Xonji 2009: 308)

(20) (a) *ketāb* *{a* *Hasan}* *om=da*
book PREP PN AC.1SG=give.PST
(b) *ketāb* *m={a* *Hasan}* *da*
book AC.1SG=PREP PN give.PST
'I gave the book to Hasan.'[22] (Xonji 2009: 256)

(21) *medād-om* *{az* *le* *zemi}* *om=vāsest-est-u*
pencil-PC.POSS.1SG PREP PREP earth AC.1SG=pick.up.PST-PP-be.PST.3SG.CRUSH
medād-om *{az* *le* *zemi}=m* *vāsest-est-u*
pencil-PC.POSS.1SG PREP PREP earth=AC.1SG pick.up.PST-PP-be.PST.3SG.CRUSH
medād-om *m={az* *le* *zemi}* *vāsest-est-u*
pencil-PC.POSS.1SG AC.1SG=PREP PREP earth pick.up.PST-PP-be.PST.3SG.CRUSH
'I had picked my pencil from the floor.'

(22) *Xiňǰ* *gāhi=m* *ne-ded-e*
PN never-AC.1SG NEG-see.PST-PP
'I have never seen Khonj.' (Xonji 2009: 308)

(23) *Hasan* *bori=m* *zat*
PN much=AC.1SG hit.PST
'I beat$_{PST}$ Hasan hard.' (Xonji 2009: 308)

17. System of Tenses: Indicative

The verbal system of Lārestāni is characterized (Skjærvø 1989: 367) as a symmetrical system of four simple tenses and corresponding continuous/progressive tenses: present-future ~ continuous present; preterit ~ imperfect; perfect ~ continuous perfect; and pluperfect ~ continuous pluperfect. In addition to these, Khonji data display a fifth indicative pair, designated here as "Past-Pluperfect." Moreover, a Past Progressive tense is discernable (see paragraph (6) below), although with scant examples. All these

[22] Note that in *m=a Hasan* the clitic is hosted by a preposition without being its object, unlike in *m=a* 'to myself'.

tenses, as exemplified in Table 3, appear in simple verb forms, i.e. single words, although certain tenses have underlying phrases, with components analyzed in §14. See also §24 for emerging periphrasis under the influence of Persian.

The semantic range of tenses seems generally compatible with those in Persian, except that the present-future and the imperfect also function as progressive tenses (§24). Nevertheless, ambiguities remain, highlighting the need for more text documentation to examine the distribution of some intricate forms in natural speech.

(1) Present-Future (or present indicative, with future and habitual functions, as in Persian) and Present Progressive (based on the past stem plus *ā*; §14.5): *a-č-eš* 'you$_{SG}$ (will) go' (Pers. miravi) ~ *čed-ā-š* (Pers. dāri miravi) 'you$_{SG}$ are going'. Despite employing the past stem, the present progressive has a nonergative alignment with transitive verbs: *ded-ā-š* 'you are seeing'.

(2) Preterit (simple past; unmarked) and Imperfect (marked durative *a-*): *čed-eš* 'you went' (Pers. rafti) ~ *a-čed-eš* (Pers. mirafti, dāshti mirafti) 'you used to go, you would go, you were going'.

(3) Perfect (present perfect; with past participle formant *-est-*) and Perfect Progressive: *čest-eš* (Pers. raftai) 'you have gone' ~ *a-čest-eš* (Pers. miraftai) 'you have been going'.

(4) Pluperfect (past perfect; with past participle in *-est-* and past copula stem *od-*) and Pluperfect Progressive: *čest_od-eš* (Salāmi *čess_ud-e*) (Pers. rafta budi) 'you had gone' ~ *a-čest_od-eš* (Pers. mirafta budi) 'you had been going' (hypothetical form; no data for intransitive verbs).

(5) Past-Pluperfect (or Perfect-Pluperfect; with past participle *-ez-* (< -est-) and perfect copula stem *bost-*) and Past-Pluperfect Progressive: *čez_bost-eš* (Pers. rafta budai) ~ *a-čez_bost-eš* (Pers. mirafta budai). According to L. Xonji, these forms have limited usage.

(6) Past Progressive. This tense, which falls outside of the five-fold symmetrical paradigm presented above, is presented by L. Xonji, with only a few examples, including *čed-āst-od-om* 'I was going', *xeled-āst-od-om* 'I was shopping' (but no transitive example). A realistic function of this form seems to be with stative verbs (§21): *od-āast-od-m* (ex. 27), 3sg. *od-āst-u* (ex. 28).

18. System of Tenses: Subjunctive

The subjunctive mood in Khonji is less commonly used than the indicative mood. Its functions are not always straightforward to identify, especially when morphologically deprived Persian is the source language in elicitations. A detailed study of the morphosyntactic structures and semantic fields of the subjunctive in Khonji, and in any other Iranian languages for that matter, can only be conducted when a sufficient amount of data based on natural speech is recorded. All I can offer here is the following classification of non-indicative moods inferred from the limited data in Khonji.

(1) Present (with the modal prefix *bé-* or preverbs):*be-ben-eš* (Pers. bebini) 'that you$_{SG}$ see'; sg. *be-ben,* pl. *be-ben-i* 'see!'; *o-č-eš* 'that you go' (cf. *a-č-eš* 'you (will) go'); sg. *očo*, pl. *oči* 'go!'. Note the irregular stem *ber-eš* 'that you come', sg. *bedā* (Salāmi *beδo*), pl. *beri* 'come!'.

(2) Progressive (with *-āz-b-* < *-āst* + *b-*, subjunctive/imperative copula): *čed-āz_beš* 'you may be going'; *čed-āz_baš* 'keep going!', *gāhi ḡossa ma-xar-dāz_baš* 'never be grieving!'. There are no parallel forms in modern Persian.

(3) Perfect (with *-ez-b-* < past part. -est + *b-*): *čedez_beš* (Pers. rafta bāši) 'you may have gone', *xatez_beš* 'you may have slept'.

(4) Pluperfect (with *boz-* < bost b-, past subjunctive copula): *čez-boz_beš* (Pers. rafta buda bāši).

19. Stress

The following stress patterns are discerned from L. Xonji and further elucidation. The stress is repelled by the durative marker *a(d)-*, pronominal clitics, and auxiliary 'be' (*-u, -e, -be, -od-, -bode*). The stress falls on the last syllable of the stem in past transitive forms (examples are in the first person singular): preterit *om=xelí* (buy); perfect *om=xeléd-e* (buy); plup. *om=soxenád-est-u* (burn); past-plup. *om=bést-ez_bode* (throw); subj. perfect *om=vāsést-ez_be* (seize) — on personal endings in the present-future: *a-nes-óm* (put) — on the infixed formant in pres. prog. *vāgašt-ā́-m* 'I am returning'; past prog. *xeled-ā́st_odom* 'I was buying'; subj. prog. *xond-ā́z_bom* 'I may be reading' — on the subjunctive morpheme: *bé-kenom* (do), *(h)ó-xatom* (sleep), but *o-čóm* (go).

Other forms show inconsistency, especially in 'come' and 'go', as in (first person singular) preterit *sótom* (burn), *oftádom* (fall) versus *ondóm* (come), *čedóm* (go); perfect *sótestom* (burn), *oftádestom* (fall) versus *ondoéstom* (come), *čéstom* (go); subj. perfect *xátez_bom* (sleep), *bódez_bom* versus *ondéz_bom* (come), *čéz_bom* (go).

Table 3. Verb forms (1sg.)[23]

	'Come'	'See'
Pres.-Future	*ad-i-om*	*a-ben-om*
Pres. Prog.	*ond-ā-(o)m*	*ded-ā-(o)m*
Preterit	*ond-om*	*om=di*
Imperfect	*ad-ānd-om*	*m=a-di*
Past Prog.	*ond-āst-od-om*	–
Perfect	*ond-est-om*	*om=ded-e*
Perfect Prog.	*ad-ānd-est-om*	*m=a-ded-e*
Pluperfect	*ond-est-od-om*	*om=ded-est-u*
Plup. Prog.	**ad-ānd-est-od-om*	*m=a-ded-est-u*
Past-Plup.	*ond-ez-bost-om*	*om=ded-ez-bode*
Past-Plup. Prog.	*ad-ānd-ez-bost-om*	*m=a-ded-ez-bode*
Subj. Pres.	*ber-om*	*be-ben-om*
Subj. Prog.	*ond-āz-b-om*	**ded-āz-b-om*
Subj. Perfect	*ond-ez-b-om*	*om=ded-ez-be*
Subj. Plup.	*ond-ez-boz-b-om*	*om=ded-ez-boz-be*

Table 4. Verb forms for 'see'

	3sg. ('he sees', etc.)	3sg. agent, 1sg. patient ('he sees me' etc.)
Pres.-Future	*a-ben-e*	*m=a-ben-e*
Pres. Prog.	*dedā-e*	*om=dedā-e*
Preterit	*oš=di*	*oš=ded-om*
Imperfect	*š=a-di*	*š=a-ded-om*
Perfect	*oš=dede*	*oš=dedest-om*
Pluperfect	*oš=dedest-u*	*oš=dedest-od-om*
Past-Plup.	*oš=dedez-bode*	*oš=dedez-bost-om*
Subj. Perfect	*oš=dedez-be*	*oš=dedez-b-om*

20. Negation

The prohibitive prefix, *ma-*, replaces the imperative markers: *biā* 'bring!', neg. *mayā*; *(hā)de* 'give!', neg. *made*; (irregular) *bedā* 'come!', neg. *mayā* (Salāmi *beδo*, neg. *mate*).

The negative prefix, *na-*, combines with the durative marker into *nā-*. Examples: *ne-nd-eš,* neg. of *ond-eš* 'you came'; *nābenom*, neg. of *a-ben-om* 'I see'; *nādānd-et*, neg. of *ad-ānd-et* 'they were coming'; *mo=ne-di*, neg. of *mo=di* 'we saw'; *mo=nādi*, neg. of *m-ā-di* 'we were seeing'.

[23] The asterisk indicates reconstructed forms missing in the data for these specific verbs.

A subdialect of Khonji employs *ya-/yo-* instead of *nă-*, as in *yo-don-om* for *nā-don-om* 'I don't know'; *ya-xel-em* for *nā-xel-em* 'we won't buy'.

A subdialect of Khonji employs *ya-/yo-* instead of *nă-*, as in *yo-don-om* for *nā-don-om* 'I don't know'; *ya-xel-em* for *nā-xel-em* 'we won't buy'.

(24) *tāvessu* *ya-be* *levās-e* *garm* *be-piš-e*
summer NEG-must? dress-EZ warm SBJV-wear.PR-2SG
'You$_{SG}$ shouldn't wear warm clothes in summer.' (Salāmi 2007: 347)

21. Stative Verbs

The progressive forms are best exemplified in the stative sense of posture verbs, i.e., verbs that appear in two senses: dynamic, involving a punctual or inchoative action; stative, involving a situation that is static or unchanging throughout its entire duration. In Khonji, the verbs 'sit', 'stand', and 'sleep', among others, have tense-differential in their dynamic and stative senses, thereby comparable with "cardinal posture verbs" in English (Newman 2009). Khonji differentiates, as does English, between the perfect and the progressive in posture verbs, whereas Persian perfect forms bear the stative sense as well. This trait is shown in Table 5 and examples that follow, for the verb 'sit', with distinct past stems: dynamic *šast-* and stative *(h)od-*. Note that the preterit has a single sense in both languages: Khonji *šast-om* ~ Pers. nešast-am 'I sat down'.

Table 5. The posture verb 'sit' (1sg.)

Sense	Tense	Khonji	Persian
dynamic	perfect	*šastest-om*	nešasta-am
stative	pres. prog.	*(h)od-ā-m*	
dynamic	pluperfect	*šastest_odom*	nešasta budam
stative	past prog.	*(h)od-āst_odom*	
dynamic	subj. perfect	*šastez_bom*	nešasta bāšam
stative	subj. prog.	*(h)od-āz_bom*	

(25) *hezār* *dafa* *ekā* *šast-est-om,*
1000 CLF here sit.PST-PP-1SG
(perfect)
ammā *āla* *ekā* *ne-hod-ā-m*
but now here NEG-sit.PST-*Ā*-1SG
(pres. prog.)
'I have sat down here a thousand times, but I am not sitting here now.' (Xonji 2009: 115)

(26) *... le* *me* *nimkat-o* *šast-est_od-om*
PREP that bench-DEIC sit.PST-PP_be.PST-1SG (plup.)
'I had sat down / taken a seat on that bench [many times].' (Xonji 2009: 126)

(27) *modir* *ke* *vāred* *bu,*
principal SUB entering it.became
mo *le* *korsi* *od-āst_od-om*
I PREP chair sit.PST-*ĀST*_be.PST-1SG (past prog.)
'When the principal entered, I was sitting on a chair.' (Xonji 2009: 232)

(28) *vaqti-ke* *Hasan* *a* *kāfa* *vāred* *bu,* *J̌amšid*
when-SUB PN PREP café entering it.became PN

le *me* *korse-n-o* *od-āst-u*
PREP that chair-EPEN-DEIC sit.PST-*ĀST* -be.PST.3SG.CRUCH (past prog.)
'When Hasan entered the coffeeshop, Jamshid was sitting [there] on that chair.' (Xonji 2009: 157)

(29) *momken-e* *sad* *dafa* *le* *e* *nimkat-o*
likely-is 100 CLF PREP this bench-DEIC

šast-ez_b-om, *ammā* *āla* *om=nā-ve*
sit.PST-PP_be.SBJV-1SG but now PC.XPER.1SG=NEG.DUR-want.PR
(subj. perfect)

ekā *od-āz_b-om;* *bāyad* *orost-om*
here sit.PST-*ĀST*_be.SBJV-1SG must rise.PR-1SG
(subj. prog.)

'It is possible that I may have sat down on this bench a hundred times, but I don't want to be sitting here now; I should stand up.' (Xonji 2009: 138)

22. Copulas

The verb 'be' has the stems *(h)-* (present), *b-* (subjunctive), *(h)od-* (past),[24] and *bost-* (perfect), conjugated with personal endings. See Table 6. Imperatives are sg. *bǎš*, pl. *bi*. The third person singular clitic *-e* is realized as *-a* after mid and high vowels: cf. *e xāla-e* 'this is the maternal aunt', *e xāle-a* 'this is the maternal uncle', *hāl-ot čʊ-a* 'how are you?', *on genā-e* 'he is crazy', *me espid-e* 'that is white'. Negatives are *nehet/nehodet* 'they are/were not'. Past copulas are used in the pluperfect.

The locative/existential verb (pres. 3sg. *ha, háste*, neg. *niste*) combines with pronominal clitics to denote possession (§23).

Table 6. The verb 'be'

	Present	Subjunctive	Preterit	Perfect
sg. 1	*(h)-om*	*bom*	*(h)odom*	*bostom*
2	*(h)-eš*	*beš, bey*	*(h)odeš*	*bosteš*
3	*(h)-e, -a*	*be*	*(h)od, -u*	*bode*
pl. 1	*(h)-em*	*bem*	*(h)odem*	*bostem*
2	*(h)-i*	*bi*	*(h)odi*	*bosti*
3	*(h)-et*	*bet*	*(h)odet*	*bostet*

'Become' is identical with 'be' in the perfect. Other tenses are formed regularly (examples in the first and third persons singular): pres.-future: *a-bó-m, abe* (Pers. mišavam); preterit: *bodom, bu* (Pers. šodam); past prog.: 3sg *abu* (Pers. mišod); perfect prog.: *abostom*, *abode* (Pers. mišoda-am); subjunctive: *bobom, bebe* (merging with 'be' in neg. *nebom*; Pers. nabāšam, našavam); imperative: sg. *bebāš*, pl. *bebi*; subj. perfect: *bodezbom* (Pers. šoda bāšam; shared with 'be'). Periphrastic passive is formed with 'become', e.g., *goto abe* 'it is said', *goto bu* 'it was said', *goto bode* 'it has been said'.

23. Possession

There is no verb 'have' in Khonji. Possession is expressed in two ways, both employing experiencer/locative constructions involving the third person singular copula with pronominal clitics as person markers.

[24] Note that *(h)od-* is also the past stem of 'sit' in its stative sense (§21). If they are cognate, the past copula may have evolved from a semantic shift from an original locative-existential meaning, signifying 'lie', 'rest'.

(a) Possession is generally expressed by pronominal clitics and the third person singular of the verb 'be': (present) *om=ha*, *otha, ošha, moha, toha, šoha,* alternatively, *omháste*, etc., neg. *omni/omniste*; (past) *om=hod, othod, ošhod, mohod, tohod, šohod*, neg. *omnehod*. Examples:

(30) *J̌amšid* *panǰ* *pos* *oš=hod*
PN 5 son PC.XPER.3SG=be.PST.3SG
'Jamshid had five sons.'[25]

(31) *balki* *en* *ketāb-o=om* *be*
maybe this book-DEIC=PC.XPER.1SG be.SBJV.3SG
'Maybe I have this book.' (Salāmi 2007: 341)

(b) Temporary possession, 'to be having, to have with self', is expressed by the base *bā́re*, as in *om= bā́re* 'I have', etc. Apparently *bā́re* consists of *bār* 'load, belongings' + 3sg. copula *-e*.[26] The preterit takes the pluperfect form: *om=bā́restu*, etc.[27] Compare:[28]

(32) – *pul=ot ha?* 'Do you have money?'

– *na, varšekast bostom.* 'No, I am broke.'

(33) – *pul=ot bāre?* 'Do you have cash on you?'

– *na, kif-e pul-om te xuna ǰā_m=nade.* 'No, I have left my wallet at home.'

24. Emerging Progressives

Progressive tenses in Khonji typically manifest through simple verb forms — present-future and imperfect (§17). This absence of periphrastic forms is emphasized by the native speaker Xonji (2009: 147-149) vis-à-vis Persian and English use of auxiliary verbs.

[25] Elicited from L. Xonji.

[26] I owe this analysis to the erudite review of this paper. Cf. *pūl-ot bā* 'you have money' in Kamioka et al. (1986: 24), where *bā* can be a short form of *bā́re*. Lāri, too, has both forms (ibid).

[27] Xonji 1999: 228.

[28] Xonji 1999: 177-178.

As if the dedicated form has grown too weak to express continuity of the action, auxiliaries are invoked: One is the particle *ase* (otherwise a preposition, §8), added optionally to reinforce progression: *(ase) Ali ded-ā-m* 'I am seeing Ali'.[29]

Moreover, Salāmi's data (2007: 342 f.) reveal three distinct Khonji constructions in response to Persian inquiries: (a) Simple verb forms, in agreement with Xonji's data, as in *šun=a-ke* 'they were doing' ~ 'they would do' ~ 'they used to do' in ex. 34. (b) Periphrasis using the spurious stem *dār-*, the present stem of Pers. dāštan 'to have', serving as an auxiliary — calquing Persian progressive tenses. Notably, the Khonji auxiliary introduces an experiencer aspect, and the preterit adopts the pluperfect form: *om=dār-essu*[30] in ex. 35 (c) Periphrasis employing *bār-*, the base accompanied by pronominal clitics to denote possession in Khonji (§23), as in ex. 36.

(34) *vakti* *ke* *seyl* *uma,*
when SUB flood come.PST.3SG.CRUSH
onyā *čekār-i_* *šun=a-ke*
they what-INDF AC.3PL=DUR-do.PST.SHORT
'What were they doing when it flooded?' (Salāmi 2007: 342)

(35) *om=dār-ess-u* *xiār* *lit* *m=a-ke,*
PC.XPER.1SG=have.Pers.-PP-be.PST.3SG.CRUSH melon slice AC.1SG=DUR-do.PST.SHORT

ke *angošt-om* *om=boli*
when finger-PC.POSS.1SG AC.1SG-cut.PST.SHORT
'I was cutting a melon, when I cut my finger.' (Salāmi 2007: 342)

(36) *om=bār-e* *levās-om* *vā-bar* *kerd-ā -m*
PC.XPER.1SG=load-be.PR.3SG dress-PC.POSS PRV-side do.PST-*Ā*-1SG
'I am putting on my clothes.' (Salāmi 2007: 343)

It should be evident that the compounds in sentences (35) and (36) share the same structure. Both must be borrowed recently from Persian (even Persian forms are relatively recent and not fully integrated into formal speech), which has also contaminated other living Iranian languages in this respect. The comparison of the two datasets (Xonji and Salāmi) highlights the value of older data in tracing language evolution. While these compounds can be emerging progressive forms, their authenticity remains in question. It is plausible that

[29] Xonji 2009: 45.
[30] Cf. *om=bā́r-estu* in §23.

the three sentences documented by Salāmi, appearing consecutively, likely from the same informant, were influenced by an elicitation method prompting the speaker to invent forms. This underscores the need for further fieldwork to ascertain whether periphrastic forms genuinely appear in natural speech beyond elicited examples.

25. Modal Forms

Constructions with the stems *(a)vest-* 'want, must' and *šā-/ša-* 'can, must, want'[31] are structured with the pronominal clitics acting as experiencers. The main verb is subjunctive if specific to a person and infinitive if general (ex. 41, 42). Both modal verbs exhibit complex conjugations, which study is beyond the scope of this paper. Here, I aim to explain the forms as illustrated in the examples below.

The verb *avesta* 'to want' (< Mid. Pers. abāyistan) has the possible present stem *(a)ve-* (< abāy-), which occurs only in the negative, e.g. *om=nāve* (ex. 37) (← na-ave or na-a-ve?), although it may be a truncated form of the past stem *(a)vest-*. The present merges with the present progressive by taking the morpheme *-ā-* (§14.5), with an ergative alignment, as in *m=avest-ā-e* or *m=a-vest-ā-e* 'I want,' literally, 'for me there is desire' (see also ex. 38). Otherwise, ergativity does not apply to present tenses regardless of transitivity (Table 3). The preterite merges with the imperfect in taking the durative marker *-a-* (ex. 38). The past participle appears as *vez-* (instead of the expected *vest-ez-; cf. *best-ez-* 'throw'), on which the subjunctive present is built; thus, 1sg. *m-avez-be*,[32] with plural clitics in longer, durative forms, as *šā-* in ex. 39 (← šo-a-vez- or šo-avez?).

(37) *om=nāve* *ke* *taxassos* *be-ger-om*
PC.XPER.1SG=NEG-want.PR SUB specialty SBJV-get-1SG
'I don't want to get a specialty.' (Xonji 2009: 317)

(38) *m=a-vest* *be-don-om*
PC.XPER.1SG=DUR-want.PST SBJV-know.PR-1SG

ammā *āla* *om=ne-vest-ā-e*
but now PC.XPER.1SG-NEG-want.PR-*Ā*-be.PR.3SG
'I wanted to know, but now I have no desire.' (Xonji 2009: 215)

[31] See also Dabir-Moqaddam 2014: §5.3.13.

[32] Corresponding with Pers. xwāsta bāšam.

(39) *mardom-e* *Kābol…* *aga* *šā=vez-be*
people-EZ PN if PC.XPER.3PL.DUR=want.PP-be.SBJV.3SG

ke *a* *Samarqand* *o-č-et…*
SUB PREP PN SBJV-go.PR-3PL
'If the people of Kabul want to go to Samarkand…' (Xonji 2009: 319)

(40) *del-om* *ša-y* *ke* *engo* *vā-mān-eš*
heart-1SG.POSS must-3SG SUB here PRV-stay.PR-2SG
'my heart desires that you stay here' (Salāmi 2007: 350)

(41) *me* *morḡ-o* *nā-šā* *xarda*
that hen-DEIC NEG.DUR-must eat.INF
'One shouldn't/can't eat that chicken.' (Xonji 2009: 223)

(42) *šomā* *tu=ya-šā* *en* *kār-o* *be-hel-i*
you.PL PC.XPER.2PL.NEG-can this work-DEIC SBJV-put.PR.-2PL
'You cannot do this work.' (Salāmi 2007: 340)

26. Lexicon

Compared to the heavily Arabicized formal Persian, the languages of southern Persia, including Lārestāni, have preserved a wealth of native words.[33] Although a comprehensive lexis of Khonji is beyond the scope of this study, a selection of Khonji words is listed below.

bač-basso 'miscarriage', *berozo* 'oven', *beu* 'bride', *bal* 'soil', *bard* 'stone', *babz* 'wasp', *bori* 'much', *buǰ* 'cork', *da(y)i, nana* 'mother', *dal* 'sparrow', *dāmu* 'maternal aunt', *dezax* 'hell', *(du)doma* 'hood, ceiling vent', *gahdim* 'north', *gerā* 'blaze', *helenǰak* 'swinging rope', *ǰā* 'room', *ǰōxan* 'stone mortar', *ǰunšur* 'bathhouse', *kala* 'hole (in walls)', *kalāt* 'fort', *kap* 'mouth', *kapferāxe* 'yawn', *kem* 'funnel', *key* (< *kas-) 'small', *kok* 'cough', *maš* 'fly', *maz* 'bee', *moh* 'palm', *mol* 'neck', *neyčit* 'straw mat', *nezbā* 'mist', *nāvā* 'waterway', *omǰal* 'cowife', *ōsu* 'winnowing fork', *oškom-ravešt* 'diarrhea', *pah* 'goat', *pahmezak* (Pers. bozmaǰa) 'lizard', *peleta* 'spark', *peva* (Pers. giǰgāh) 'temple', *pinǰa* 'finger', *pop* 'lung', *rāvand* 'coffin', *sangara* 'ice', *sendu* 'constipation', *songe, sag* 'dog', *sur, ser* 'salty', *šādi* 'monkey', *šāt* 'wax', *šek* 'owl', *taš* 'fire', *tela* 'newborn', *telaza* (Pers. zāʾu) 'puerpera', *xaǰa* 'firewood', *xars* 'tears', *xarčo* 'gutter', *xāg* 'egg', *xāla* 'maternal aunt', *xāle* 'maternal uncle', *xog* 'corner'.

[33] See Borjian 2019: §4.1.

ABBREVIATIONS

:	long vowel
:	separates present and past stems
_	phonological linker between words
-	morpheme separator
=	separates agent clitics
∅	zero morpheme
1, 2, 3	first, second, third person-agreement marker
-Ā-	present progressive marker (§14.5)
AC	agent clitic (§16)
-ĀST-	past and subjunctive progressive marker *-āst-, -āz-* (§14.6)
CAUS	causative (§12)
CLF	numeral classifier
COP	copula (Table 6)
CRUSH	crushed stem (third person singular) (§15)
DEF	definite *-é* (§4)
DEIC	deictic *-o* (§4)
DIMIN	diminutive (§4)
DUR	imperfective *a(d)-* (§14.2)
EPEN	epenthesis
EZ	ezafe (§4)
ex.	exampled sentence
INDF	indefinite
INF	infinitive
INTR	intransitive
NEG	negation, negative (§20)
OBJ	object
PC	pronominal clitic (Table 1)
Pers.	(modern) Persian
pl., PL	plural
PLUP	pluperfect
PN	proper noun
POSS	possessive (§6.1)
PP	past participle formant *-est-, -ez-, -e* (§14.3-4)
PREP	preposition (§8)
pres., PR	present
PRFCT	perfect
PRV	preverb (§13)
PST	past
REFL	reflexive (§6.7)
SBJ	subject

SBJV	subjunctive/imperative (§18)
sg., SG	singular
SHORT	shortened or truncated stem (§11)
SUB	subordinator
TR	transitive
V	any vowel
XPER	experiencer (§23, §25)

BIBLIOGRAPHY

Barbera, G. (2005), "Lingua e cultura a Minâb (Iran Sud-Orientale). Profilo grammaticale, testi e vocabolario," doctoral diss., Naples Eastern University, 2005.

Borjian, H. (2017), "Kerman xvi. Languages." *Encyclopaedia Iranica* XVI/3. 290-304.

Borjian, H. (2019), "The Language of the Kharg Island." *Journal of the Royal Asiatic Society*, 29/4. 659-682.

Borjian, H. (2020), "The Perside Language of Shiraz Jewry: A Historical-Comparative Phonology." *Iranian Studies* 53/3-4. 403-415.

Borjian, H. (forthcoming), *The Perside Language of Shiraz Jewry. Grammar, Texts, Lexicon*.

Calmard, J. (1986), "Lār, Lāristān." *The Encyclopaedia of Islam, New ed.*, vol. 5, E. J. Brill, pp. 665-676.

Dabir-Moqaddam, M. (2014), *Radašenāsi-e zabānhā-ye irāni*, 2 vols., Tehran: Samt.

Eqtedāri, A. (1955), *Farhang-e lārestāni.* Tehran: Farhang-e Irānzamin.

Gershevitch, I. (1970), "The Crushing of the Third Singular Present." In M. Boyce and I. Gershevitch (eds.), *W. B. Henning Memorial Volume*, 161-74. London: Lund Humphries.

Haïm, S. (2001), *The Larger Persian-English Dictionary*, 16th edition, Tehran: Farhang.

Kalbāsi, I. (2009), *Farhang-e towṣifi-e gunahā-ye zabāni-e Irān*. Tehran: Pažuheśgāh-e ʿolum-e ensāni.

Kamioka, K., A. Rahbar, and A. A. Hamidi (1986), *Lārestanī Studies* II. *Comparative Basic Vocabulary of Khonjī and Lārī*. Tokyo: Institute for the Study of Languages and Cultures of Asia and Africa.

Molčanova, E. K. (1982), "Larskij jazyk." In *Osnovy iranskogo jazykoznanija. Novoiranskie jazyki: Zapadnaya gruppa, Prikaspijnskie jazyki*, 364-446. Moscow: Nauka.

Molčanova, E. K. (1997), "Larskaja gruppa dialektov." In *Jazyki mira. Iranskie*

jazyki I. *Jugo-zapadnje iranskie jazyki*, 174-177. Moscow: Indrik.

Newman, J. (2009), "English posture verbs: An experientially grounded approach," *Annual review of Cognitive Linguistics* 7. 30-57.

Ṣādeqi, ʿA.-A. (2003), "Darbāra-ye guyeš-e lārestāni." *Majalla-ye Zabānšenāsi* 18.1 (no. 35). 120-138.

Salāmi, ʿA.-N. (2007), *Ganjina-ye guyeššenāsi-e Fārs* IV. Tehran: Farhangestān.

Skjærvø, P. O. (1989), "Languages of Southeast Iran: Lārestānī, Kumzārī, Baškardī." In R. Schmitt. (ed.), *Compendium Linguarum Iranicarum*, 363-69. Wiesbaden: Reichert.

Vosuqi, M. B. (1995), *Xonj, gozargāh-e bāstāni-e Lārestān*. Qom: Xorram.

Xonji, A.-Ḥ. (2015), *Guyeš-e lārestāni*. Shiraz: Navid.

Xonji, L.-ʿA. (1999), *Dastur-zabān-e lārestāni bar mabnā-ye guyeš-e xonji*. Shiraz: Ilāf.

Xonji, L.-ʿA. (2009), *Negareš-i tafṣili bar zabān-e lārestāni o guyeš-e xonji*. Shiraz: Dānešnāma-ye Fārs.

YEREVAN STATE UNIVERSITY PUBLISHING HOUSE

JIL 1 (2024): 118-141

Journal of Iranian Linguistics

EZAFE IN THE CONTEXT OF CPS: Evidence from three Iranian languages

Songül Gündoğdu
Muş Alparslan University

Arsalan Kahnemuyipour
University of Toronto Mississauga

Marcel den Dikken
Eötvös Loránd University
Hungarian Research Centre for Linguistics

doi.org/10.46991/jil/2024.01.06

Abstract: The present study investigates the distribution of the Ezafe (EZ) morpheme in adnominal clauses in three Iranian languages, namely Persian, Kurmanji (Northern Kurdish), and Zazaki, demonstrating that the behavior of EZ in these languages challenges the case analysis of EZ, suggesting instead a compatibility with the inversion analysis of EZ. In contrast to the prediction made by the case analysis, it is shown that EZ is required before [–N] modifiers such as CPs. First, EZ is consistently used in restrictive relative clauses (RCs) in all three languages, with Persian using an allomorph of EZ in this context, contrasting with Kurmanji and Zazaki, which use the regular form of EZ. Non-restrictive RCs present divergence: while Persian does not allow EZ in this environment, Zazaki and Kurmanji employ regular and anaphoric EZ (AEZ) forms, respectively. Following de Vries (2006), non-restrictive RCs are treated as restrictive RCs with a silent head, aligning the distribution of EZ in these languages with its distribution after a silent noun: Persian lacks EZ here, while Zazaki and Kurmanji use EZ and AEZ, respectively. Second, in Noun-Complement Clauses (NCCs), Kurmanji and Zazaki consistently use regular EZ, while Persian offers two options: the allomorph of EZ used with CPs or no EZ at all. Two possible structures are proposed for NCCs, one with and one without inversion, attributing the distribution of EZ to the structure involving inversion. This study shows that EZ is present in CP contexts and its distribution in these languages follows from the general behaviour of EZ and the syntax of N-CP structures.

Keywords: Ezafe, Iranian languages, relative clauses (RCs), noun-complement clauses (NCCs), CPs, silent noun.

Songül Gündoğdu
E-mail:
s.gundogdu@alparslan.edu.tr

Arsalan Kahnemuyipour
E-mail:
a.kahnemuyipour@utoronto.ca

Marcel de Dikken
E-mail: dmarcel@nytud.hu

Received: 15.03.2024
Revised: 10.07.2024
Accepted: 20.07.2024

This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License

Introduction

1. Introduction *

In many Iranian languages, a linking element known as 'Ezafe' (hereafter EZ) appears between a noun and its postnominal modifier, including possessors, and is repeated on subsequent modifiers, if they are present, except the last one (Samiian 1994; Ghomeshi 1997; Samvelian 2007; Larson and Yamakido 2008; Haig 2011; Kahnemuyipour 2014; Larson and Samiian 2020, 2021; among others), as indicated in the following schema with multiple modifiers:

(1) N-**Ez** MOD_1-**Ez** MOD_2-**Ez** MOD_3

The form of the Ezafe morpheme is invariant (barring minor phonological modifications) in Persian, where it appears as *-e* (or *-ye* after a vowel), as illustrated in (2).

(2) Ezafe in Persian

a. *(ye)* *boz-e* *siāh*
a goat-EZ black
'a/the black goat'

b. *(ye)* *mard-e* *čāq*
a man-EZ fat
'a/the fat man'

c. *sib-e* *qermez-e* *bozorg*
apple-EZ red-EZ big
'red big apple'

d. *ketāb-e* *Ali/man*
book-EZ Ali/1SG
'Ali's/my book'

* We gratefully acknowledge funding from the Social Sciences and Humanities Research Council of Canada (SSHRC Grant #435-2018-0527). Some of the work on this paper was completed when the first author held the Elahe Omidyar Mir-Djalali Postdoctoral Fellowship in Iranian Linguistics at the University of Toronto, generously funded by Roshan Cultural Heritage Institute and the aforementioned SSHRC grant. We would like to thank the members of the "Syntax of Nominal Linkers" project and the audiences of CLA 2021, ICKL 2023 and NACIL 2023 for their feedback on earlier versions of this work. We are truly grateful to Serpil Güngör and Amine Bulan for generously sharing their knowledge of Zazaki. We also sincerely thank an anonymous reviewer for their constructive feedback. All errors are ours.

In certain Iranian languages, such as Kurmanji (Northern Kurdish), the EZ morpheme cross-references the phi-features of the head noun and definiteness, as illustrated in (3). The Kurmanji Ezafe morpheme agrees with the feminine 'goat' in (3a/3a') whereas it agrees with the masculine 'man' in (3b/3b'). Moreover, the definiteness ('the goat / the man') vs. indefiniteness ('a goat / a man') of the head noun is reflected on the Ezafe vowel, as shown by the distinction in (3a/a') and (3b/b'). Gender distinction is neutralized in the plural (3c), and phi feature agreement remains consistent in both the modifier and possessor environment (compare 3a and 3d).

(3) Ezafe in Kurmanji

a. *bizin-a* *reş*
goat-EZ.F black
'the black goat'

a'. *bizin-ek-e* *reş*
goat-INDF-EZ.F black
'a black goat'

b. *mirov-ê* *qelew*
man-EZ.M fat
'the fat man'

b'. *mirov-ek-î* *qelew*
man-INDF-EZ.M fat
'a fat man'

c. *sêv-ên* *sor-ên* *mezin*
apple-EZ.PL red-EZ.PL big
'the big red apples'

d. *kitab-a* *Ali/min*
book-EZ.F Ali/1SG.OBL
'Ali's/my book'

In another Iranian language, Zazaki, Ezafe inflects for number, gender and case. The forms in (4a/b/c) represent the Direct forms of the Ezafe morpheme. The different forms of Ezafe can be seen clearly when the modified noun is placed in various case positions. For instance, when the masculine modified noun is the subject of a transitive clause in the present tense (a DIR case position), the DIR form of the Ezafe is used. Meanwhile,

when the modified noun appears in an oblique context, e.g. the direct object of a present-tense verb, the OBL form of the Ezafe is used. We can see the contrast between the DIR and OBL forms of Ezafe in the masculine singular, *-o* and *-ê*, respectively, in (4d/e).

(4) Ezafe in Zazaki

a. *bız-a* *sıpê*
goat-EZ.F white
'the white goat'

b. *kutık-o* *sıpê*
dog-EZ.M white
'the white dog'

c. *sol-ê* *sıpê*
shoe-EZ.PL white
'the white shoes'

d. [*Kutık-o* *sıpê*] *mın*
dog-EZ.M.DIR white 1SG.OBL
vin-en-o.
see-PRS-3SG.M
'The white dog sees me.'

e. *Ez-a* [*kutık-ê*
1SG.DIR-PROG.1SG dog-EZ.M.OBL
sıpê] *vin-en-an.*
white see-PRS-1SG
'I see the white dog.'

The focus of the current study is the distribution of EZ in the context of nouns followed by CPs, both relative clauses (RCs) and so-called noun-complement clauses (NCCs) in three different Iranian languages, namely Persian, Kurmanji and Zazaki.[1]

One prominent analysis of EZ takes it to be a case assigner required before all [+N] elements (Samiian 1994; Larson and Yamakido 2008; Larson and Samiian 2020, 2021). This type of analysis predicts that adnominal elements which are [–N] should not be preceded by EZ. Persian non-restrictive RCs seem to provide support for this analysis as they are not preceded by EZ. Meanwhile, restrictive RCs are preceded by a (so-called

[1] For a more detailed description of Ezafe in these three (and several other) Iranian languages, see Taghipour and Kahnemuyipour (2023).

relative) particle *-i*, phonologically distinct from the regular EZ *-e*. This particle has been analyzed as an allomorph of EZ, which then presents a counter-example to the case analysis (Kahnemuyipour 2014). The idea that Persian *-i* before restrictive RCs is a form of Ezafe finds further support in Kurmanji and Zazaki, which use the regular form of EZ with restrictive RCs. Under this view (*contra* the case analysis), EZ is used uniformly before a modifier, regardless of its [+/−N] status. Non-restrictive RCs in Kurmanji and Zazaki add an interesting twist to the discussion, as in these contexts, both languages allow EZ. While Zazaki uses the regular form of EZ preceded by prosodic break, Kurmanji employs a different type of EZ known as anaphoric EZ (AEZ).

We argue in this paper that the distribution of EZ in the context of adnominal clauses in Kurmanji and Persian poses a serious challenge to the case analysis of EZ, which predicts that [–N] modifiers should not require the presence of EZ. We further demonstrate that the facts from these two languages are instead compatible with the inversion analysis of EZ (Kahnemuyipour 2014), given a proper understanding of the syntax of N–CP structures.

This paper is structured as follows. The next section provides a brief overview of two prominent syntactic analyses of Ezafe, namely the case analysis and the inversion analysis, and the predicted distribution of Ezafe in N–CP structures. Section 3 presents the distribution of Ezafe in the context of RCs in Persian, Kurmanji and Zazaki providing arguments against the case analysis. We argue that the distribution of EZ in the context of RCs in these languages follows from the general behavior of EZ and the syntax of relative clause structures. Section 4 discusses the distribution of Ezafe in the context of NCCs in these three Iranian languages and provides two possible structures for NCCs: (i) NCC as the subject of predication for the projection of the head noun, with the surface order derived as a result of inversion of NP around CP, and (ii) NCC as (a subpart of) the predicate for the projection of the head noun with no inversion involved. We posit that while Persian allows both strategies, Kurmanji and Zazaki employ the former only. This division corresponds to the optional or obligatory presence of a nominal linker. Section 5 presents concluding remarks with empirical and theoretical implications.

2. Two syntactic accounts of Ezafe: Case or Inversion

As a distinguishing grammatical feature of noun phrases in many Iranian languages, Ezafe has been a source of interest for theoretical linguists. Two prominent syntactic accounts of Ezafe take EZ to be either *a case assigner* (Samiian 1994; Larson and Yamakido 2008; Larson and Samiian 2020) or *a*

reflex of inversion (Kahnemuyipour 2014). The case analysis of EZ assumes a DP structure where all NP modifiers originate postnominally and as [+N] elements they need to be case-licensed. Under this view, while the head noun is case-licensed by D, all other [+N] modifiers (including adjectives, possessors, etc.) are case-licensed by EZ. Thus, for example, in (2c)-(3c) repeated here as (5a)-(5b), the first EZ case-licenses "red" and the second EZ "big".

(5) a. Persian
sib-e *qermez-e* *bozorg*
apple-EZ red-EZ big
'big red apple'

b. Kurmanji
sêv-ên *sor-ên* *mezin*
apple-EZ.PL red-EZ.PL big
'the big red apples'

The case analysis of EZ makes the prediction that [–N] modifiers should not require (or even desire) the presence of EZ. In particular, if a head noun is followed by PP or CP, no EZ should be required between them (Samvelian 2007; Kahnemuyipour 2014).[2] This study focuses on the N–CP context, investigating how this prediction of the case analysis of EZ fares with the facts in Persian, Kurmanji and Zazaki.

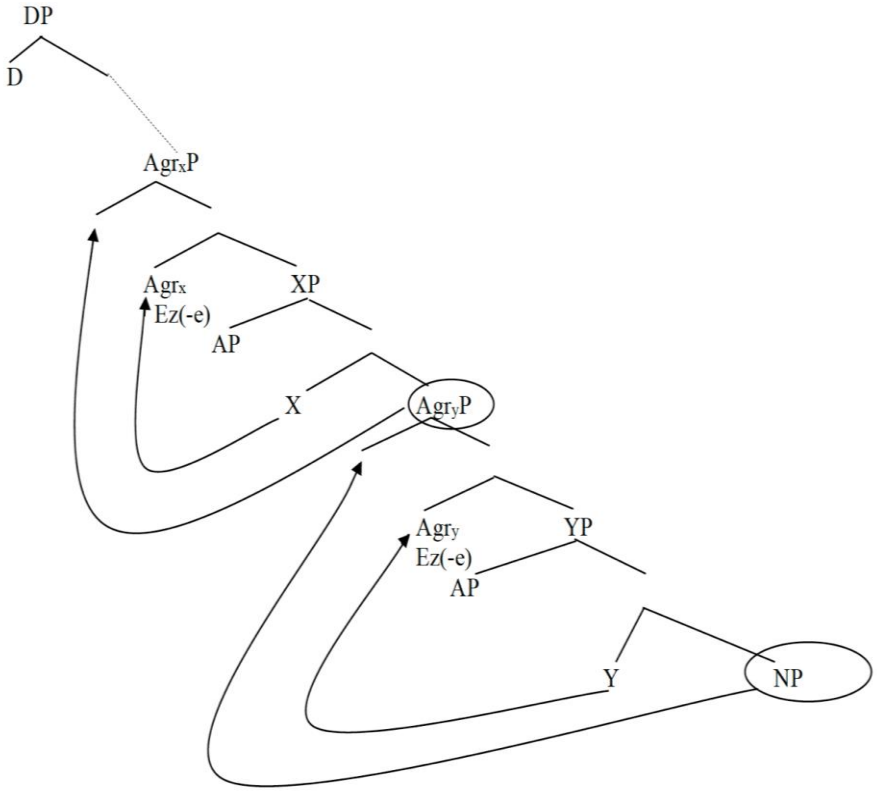
According to the inversion analysis of Ezafe, the noun phrase in Ezafe languages is taken to be head-final, with the modifiers residing in the specifiers of projections above N.[3] In addition to the projections which house the modifiers, there are intermediate projections which enable the roll-up derivation, shown schematically in the tree diagram in (6). Under this view, the Ezafe marker can be seen as the surface realization of the suggested inversion process (akin to den Dikken's 2006 linker). Crucially, this account

[2] As is well established in the literature on Persian and other Iranian languages more generally (Samiian 1994; Ghomeshi 1997; Karimi and Brame 2012; Kahnemuyipour 2014; Larson and Samiian 2021; among others), P(reposition)s are divided into two main classes, nominal Ps which take the Ezafe marker, and true Ps which do not. Accordingly, an EZ is expected between a noun and a modifying PP if the P is a nominal P and not a true P. In this paper, we are abstracting away from the N–PP context (see Samvelian 2007; Kahnemuyipour 2014; Larson and Samiian 2021, Kahnemuyipour and Taghipour 2024, for discussion).

[3] This structure is in line with other roll-up analyses of DP structure in other languages within the framework best known as cartography (Cinque 2002, 2005, 2010; Shlonsky 2004, 2010, among others).

does not predict a blanket absence of EZ in the context of N–CP, an issue we turn to immediately below.

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



In the following sections, we argue that the distribution of EZ in the context of RCs and NCCs in Persian, Kurmanji and Zazaki poses serious challenges to the case analysis of EZ, which predicts that [–N] modifiers should not require the presence of EZ, whereas the facts from these three languages are instead compatible with the inversion analysis of EZ, which predicts that EZ is used uniformly before a modifier, regardless of its [+/−N] status.

3. Ezafe in the Context of Relative Clauses (RCs)

3.1. Persian

In the previous section, we discussed how the case analysis of EZ predicts the absence of EZ in the context of [–N] modifiers of NP. In apparent accordance with this, Persian non-restrictive RCs are not preceded by EZ (7).

(7) a. *Dust-e* *Hasan,* *ke* *tu* *Tehrān*
friend-EZ Hasan that in Tehran
dars *mi-xun-e*,
lesson DUR-read.PRS-3SG
xeyli *bāhush=e.*
very smart=COP.PRS.3SG
'Hasan's friend, who is a student in Tehran, is very smart.'

b. *Xāhar-e* *man,* *ke* *tu* *Tehrān*
sister-EZ 1SG that in Tehran
dāneshju=e, *emruz*
student=COP.PRS.3SG today
mi-yā-d *injā.*
DUR-come.PRS-3SG here
'My sister, who is a student in Tehran, is coming here today.'

Meanwhile, restrictive RCs are preceded by a (so-called relative) particle *-i* (8), which is phonologically distinct from the regular EZ *-e*.

(8) a. *Zan-**i*** *ke* *az* *Tehrān*
woman-**i** that from Tehran
umad-e *xeyli*
come.PST-PERF very
bāhush=e.
smart=COP.PRS.3SG
'The woman who has come from Tehran is very smart.'

b. *Mard-**i*** *ke* *dār-am* *bāh=āsh*
man-**i** that have-1SG with=3SG
telefon-i *sohbat* *mi-kon-am*,
phone-ADJ speak DUR-do.PRS-1SG
pesar-xāl=am=e.
son-uncle=1SG=COP.PRS.3SG
'The man whom I am talking to on the phone is my cousin.'

This particle has puzzled Persian syntacticians for a long time. In Kahnemuyipour (2014), this particle is analyzed as a grammatically-conditioned allomorph of EZ.[4] In dialects of Dari (Persian spoken in Afghanistan), the same form *-i* is used as regular Ezafe, e.g. *doxtar-i xord* 'little daughter', as well as with restrictive RCs (9), cf. (8a), lending further support to the idea that the particle used with restrictive RCs in Iranian Persian is an allomorph of Ezafe (see also Samiian and Larson's 2023 discussion of their example (16b)). If this approach to *-i* is correct, and if, as standardly assumed, restrictive relative clauses are [–N], (8) and (9) undermine the case analysis of EZ.

(9) *Zan-**i*** *ke* *az* *Tehrān* *āmada*
woman-**i** that P Tehran come.PST-PERF
besyār *hushyār* *ast.*
very smart COP.PRS.3SG
'The woman who has come from Tehran is very smart.'

3.2. Kurmanji and Zazaki

Further support for the idea that EZ is at work in restrictive relativization comes from Kurmanji (10) and Zazaki (11), which use the regular form of EZ uniformly in front of any restrictive modifier (regardless of its [+/−N] status), including RCs:

(10) a. *Jin-a* *ku* *ji* *Stenbol-ê*
woman-EZ.F that P Istanbul-OBL
hat-iy-e *gelek* *zîrek* *e.*
come.PST-3SG-PERF very clever COP.PRS.3SG
'The woman who has come from Istanbul is very clever.'

b. *Mêrik-ê* *ku* *ez* *bi* *telefeon-ē*
man-EZ.M that 1SG.DIR P telephone-OBL
pê *ra* *xeber* *di-d-im*
P.3SG.OBL Prt. Info PROG-give.PRS-1SG
pismam-ê *min* *e.*
cousin-EZ.M 1SG.OBL COP.PRS.3SG
'The man whom I am talking to on the phone is my cousin.'

[4] The connection between Ezafe and the so-called relative particle finds support from a historical perspective, as the Persian Ezafe is taken to be a descendent of the Old and Middle Persian 'relative connector', used to connect the noun with the post-nominal restrictive relative clause (Samvelian 2007, Skjærvo 2009, Kahnemuyipour 2014, also Moyne and Carden 1974).

(11) a. *Cinik-a* *ki* *Istembul* *ra*
woman-EZ.F that Istanbul P
am-a *zaf* *derg* *a.*
come.PST-3SG.F very tall COP.PRS.3SG.F
'The woman who has come from Istanbul is very tall.'

b. *Merık-o* *ki* *ez* *telefon* *de*
man-EZ.M.DIR that 1SG.DIR telephone P
tey *qeşi* *ken-an,*
together with speak do.PRS-1SG
lac-ê *ap-ê* *mın*
son-EZ.M.OBL uncle-EZ.M.OBL 1SG.OBL
o.
COP.PRS.3SG.M
'The man whom I am talking to on the phone is my cousin.'

Non-restrictive RCs in Kurmanji and Zazaki add an interesting twist to the data presented above. In these contexts, both languages allow EZ. While Zazaki employs the regular form of EZ preceded by a prosodic break (12), Kurmanji uses a different type of EZ, which is known as anaphoric EZ (AEZ) (13) (Haig 2011). Anaphoric EZ is distinguished from the regular EZ by the use of an initial glide. This is in contrast to Persian, which does not use EZ at all in cases of non-restrictive relativization (7).

(12) a. *Embaz-ê* *Hesen-i,* *o* *ki* *Istanbol*
friend-EZ.M.OBL Hasan-OBL EZ.M.DIR that Istanbul
ra *ama,* *zaf* *gureci* *yo.*
P come.PST.3SG very hardworking COP.PRS.3SG.M
'Hasan's friend, who came from Istanbul, is very hardworking/capable.'

b. *Wu-a* *min,* *a* *ki* *çend* *asmiyo*
sister-EZ.F 1SG.OBL EZ.F that some month
nê-ven-en-an, *en-a* *suk-i.*
NEG-see-PRS-1SG come.PRS-3SG.F town-OBL
'My sister, whom I haven't seen in months, is coming to town.'

(13) a. *Heval-ê* *Hasan,* *yê* *ku* *li* *Stenbol-ê*
friend-EZ.M Hasan AEZ.M that P Istanbul-OBL
xwand, *gelek* *zîrek* *e.*
read.PST.3SG very clever COP.PRS.3SG
'Hasan's friend, who was a student in Istanbul, is very clever.'

b. *Xwişk-a* *min,* *ya* *ku* *çend* *meh*
sister-EZ.F 1SG.OBL AEZ.F that some month
min *ne-dît-i-ye,* *îro* *tê*
1SG.OBL NEG-see.PST-PERF-3SG today PROG.come.PRS.3SG
mal-ê.
home-OBL
'My sister, whom I haven't seen in months, is coming home today.'

3.3. Summary

The distribution of Ezafe in the context of relative clauses in these Iranian languages is summarized in the table below.

Table 1. Distribution of EZ with Relative Clauses in Persian, Kurmanji and Zazaki

	Restrictive RCs	Non-restrictive RCs
Persian	EZ (*-i*)	—
Kurmanji	EZ	AEZ
Zazaki	EZ	EZ

Once we take the so-called relative particle in Persian to be an allomorph of EZ, the distribution in the context of restrictive RCs shown in Table 1 can be understood as the regular use of EZ with modifiers more generally. In other words, EZ can be said to appear uniformly before a restrictive RC in Kurmanji, Zazaki and Persian alike, with the only difference that Persian uses an allomorph of EZ in this context.

3.4. The analysis of non-restrictive relativization and Ezafe

For the syntax of non-restrictive relativization, we follow de Vries (2006), who proposes that the relative clause is a restrictive modifier of a noun phrase headed by a silent noun or nominal proform. Under this view, the relativized noun phrase specifies the content of the projection of the physical head noun, and is connected to it via asyndetic coordination, established in the 'colon phrase' in the structure in (14a).

(14) a. [$_{:P}$ [$_{DP}$ *John*] [**:** [$_{DP}$ D [$_{NP}$ ONE/PERSON$_{Ø}$] [$_{CP}$ *who loves Mary*]]]]
b. *John, who loves Mary = John, viz., THE ONE/PERSON who loves Mary*

From (14), the distribution of EZ with Kurmanji and Zazaki non-restrictive RCs follows directly, as it matches the distribution of EZ following a silent N more generally: while Persian does not allow EZ in these contexts, Zazaki and Kurmanji use EZ and AEZ, respectively, (15). In other words, the presence or absence of EZ with non-restrictive RCs is not an idiosyncratic property. If a language like Persian does not allow the presence of EZ after silent Ns, no EZ will be used with non-restrictive RCs (15). If a language, such as Kurmanji or Zazaki, allows for the presence of EZ after a silent N, the same linker is used in the context of non-restrictive RCs, (16).[5]

---

[5] One might think that an approach that takes non-restrictive RCs to be enveloped in a projection of a silent head noun which is, in turn, juxtaposed to the projection of the overt head noun could introduce a novel opportunity to take EZ with non-restrictive relative clauses to be the reflex of case assignment (*à la* Larson and Samiian), if one assumes there to be a case relation between the head noun and the silent-headed NP (indubitably [+N]) that envelops the RC. Coupled with a perspective on the distribution of EZ in the context of silent nouns more generally, this could conceivably capture the relevant facts. But assuming there to be a case relation between the overt head noun and the silent-headed NP asyndetically coordinated with it would be quite problematic, for the following reasons. First, case is usually taken to be associated with overt Ns, not silent ones. Second, in other silent-N EZ contexts (see (16) as well as (i) below), no overt N precedes the silent N, making it unlikely that a case-based approach could capture all silent-N cases uniformly. Lastly, on a de Vries-style asyndetic coordination approach (on which the relationship between the projection of the overt head noun and the projection of the silent noun is one of asyndetic coordination), case assignment to the second conjunct is unexpected in light of the fact that case is not normally assigned to second conjuncts separately, let alone by or from the first conjunct.

(i) a. Kurmanji

Ez	hesp-ê	reş	na-xwaz-im,	**yê**
1SG.DIR	horse-EZ.M	black	NEG-want.PRS-1SG	AEZ.M
spî	di-xwaz-im.			
White	PROG-want.PRS-1SG			

'I don't want the black horse; I want the white one.' (Gündoğdu 2023: 13)

b. Zazaki

Ez-a	kıtab-ê		suri	nê	ê
1SG.DIR-PROG	book-EZ.M.OBL		red	not	EZ.M.OBL
kıhoyi	g-en-a.				
blue	get-PRS-1SG				

'I get the blue book not the red one.' (Gündoğdu and Bulan 2023: 15)

(15) Persian

Man	*xodkār-e*	*ābi-ro*	*bā*	*Ø$_N$(*-e) .*
1SG	pen-EZ	blue-RA	with	EZ
qermez	*avaz*	*kard-am.*		
red	change	do.PST-1SG		

'I exchanged the blue pen with a red one.'

(16) a. Kurmanji

Min	*qelem-a*	*şîn*	*bi*
1SG.OBL	pen-EZ.F	blue	with
Ø$_N$(-ya)*	*sor*	*guhart.*	
AEZ	red	change.PST.3SG	

'I exchanged the blue pen with the red one.'

b. Zazaki

Mın	*qelem-a*	*şin-ı*	*bı*	*Ø$_N$*(-a)*
1SG.OBL	pen-EZ.F	blue-OBL.F	with	EZ.F
sûr-a	*vurn-a*		*ya.*	
red-OBL.F	change.PST-3SG.F		Part	

'I exchanged the blue pen with the red one.'

So far, we have looked at the distribution of EZ in the context of RCs in Persian, Kurmanji and Zazaki and shown how it follows from the general distribution of EZ and the syntax of RCs. Next, we consider the Noun-Complement Clause context.

4. Ezafe in the Context of Noun Complement Clauses (NCCs)

Kurmanji and Zazaki NCCs are always linked to the head N with EZ (N-EZ CP) as illustrated in (17) and (18), respectively, while Persian has been claimed to lack EZ in NCC contexts (19).

(17) Kurmanji

a.

Hêvi-ya	*ku*	*Şah*	*ji*	*Îran-ê*		
hope-EZ.F	that	Shah	P	Iran-OBL		
derkev-e,		*roj*	*bi*	*roj*	*winda*	*bû.*
SBJV.go out.PRS-3SG		day	with	day	lose	COP.PST.3SG

'The hope that Shah would leave Iran faded over time.'

b. *Ew* *gotegot-a* *ku* *vaksîn/derzî* *bêkêr*
DEM.DIR rumor-EZ.F that vaccine useless
e *di-b-e* *sedem-a*
COP.PRS.3SG PROG-become.PRS-3SG reason-EZ.F
xem-an.
concern-OBL.PL
'The rumor that the vaccine is useless is causing concern.'

(18) Zazaki

a. *omıd-**o*** *ki* *şax* *İran* *ra*
hope-EZ.DIR.M that Shah Iran P
veci-yo *roc bı roc* *bi*
leave.PRS-3SG.M day by day become.PST.3SG
kemi.
Less
'The hope that Shah would leave Iran faded over time.'

b. *a* *xeber-**a*** *ki* *derjin/aşi*
DEM.F rumor/news-EZ.F that vaccine
bêfaydı *o* *sebeb-ê*
useless COP.PRS-3SG.M reason-EZ.PL
qısawat-ı *ben-a.*
problem-OBL.F SBJV.COP.PRS-3SG.F
'The rumor that the vaccine is useless is causing concern.'

(19) Persian

in *omid (*-e)* *ke* *Shāhæz* *Irān* *xahæd=ræft*
this hope -EZ that Shah from Iran will=go
'the hope that the Shah will leave Iran' (Larson and Samiian 2020: 200)

Larson and Samiian (2020) attribute this difference to the alleged [+N] status of CPs in Kurmanji, as opposed to Persian. They base their claim that CPs are [+N] in Kurmanji but [–N] in Persian on the following argument with respect to relative clauses. They suggest that Kurmanji *ku* is a relative pronoun while Persian *ke* is a complementizer based on the observation that cross-linguistically, relative clauses introduced by a complementizer allow resumptive pronouns but RCs with a relative pronoun do not. Persian allows resumption under relativization under certain circumstances while Kurmanji does not; *ergo*, Kurmanji *ku* is a relative pronoun but Persian *ke* is a complementizer.

This argument has four limitations. First, from the conclusion that Kurmanji *ku* is a relative pronoun (and as standardly assumed, in SpecCP), nothing follows regarding the specification of the relative CP for the feature [+/–N] as CPs do not 'inherit' their categorial feature content from the operator in their specifier; the external distribution of relative clauses is not determined by the categorial features of the relative operator. Second, the conclusion that Kurmanji *ku* is a relative pronoun does not straightforwardly carry over to the syntax of noun-complement clauses (though see Krapova and Cinque (2015), where NCCs are analyzed as reduced relative clauses; cf. fn. 6 below), whose presumed specification for the feature [+/–N] remains largely unsupported. Third, the distribution of Persian and Kurmanji CPs elsewhere is identical: CPs cannot be used as clausal subjects in either language without an additional nominal element, e.g. a demonstrative (20), and in both Persian and Kurmanji, CP complements are post-verbal unlike nominal arguments, which are preverbal, (21/22). Therefore, the claimed contrast between Kurmanji and Persian is not robustly supported. It is worth noting that CPs have the same distribution in Zazaki as well, as shown in (23).

(20) a. Persian

**(in)*	*ke*	*vāksan*	*bifāyde*	*ast*
this	that	vaccine	useless	COP.PRS.3SG
kāmelan	*doruq=e.*			
totally	lie=COP.PRS.3SG			

'That the vaccine is useless is totally false.'

b. Kurmanji

Ew ku	*derzî*	*bêkêr*	*e*
DEM	that vaccine	useless	COP.PRS.3SG
hemû	*derew*	*e.*	
all	lie	COP.PRS.3SG	

'That the vaccine is useless is all a lie.'

(21) Persian

a. Nominal arguments

Man	*ketāb-o*	*mi-xun-am.*
1SG	book-RA	DUR-read.PRS-1SG

'I am reading the book.'

b. CP complements
Un ne-mi-dun-e ke man ketāb-o
3SG NEG-DUR-know.PRS-3SG that SG book-RA
mi-xun-am.
DUR-read.PRS-1SG
'S/he doesn't know that I am reading the book.'

(22) Kurmanji
a. Nominal arguments
Ez kitab-ê di-xwîn-im.
1SG.DIR book-OBL PROG-read.PRS-1SG
'I am reading the book.'

b. CP complements
Ew ni-zan-e ku ez
3SG.DIR NEG-know.PRS-3SG that 1SG.DIR
kitab-ê di-xwîn-im.
book-OBL PROG-read.PRS-1SG
'S/he doesn't know that I am reading the book.'

(23) Zazaki
a. CP as a clausal subject
A ki derjin/aşi bêfaydı o
DEM.F that vaccine useless COP.PRS-3SG.M
sebeb-ê qısawat-ı ben-a.
reason-EZ.PL problem-OBL.F SBJV.COP.PRS-3SG.F
'That the vaccine is useless is causing concern.'

b. CP complements
A ne-zan-a ki ez
3SG.DIR.F NEG-know.PRS-3SG.F that 1SG.DIR
kitab-ı wan-en-an.
book-OBL.F read-PRS-1SG
'She doesn't know that I read the book'

c. Nominal arguments
Ez kitab-ı wan-en-an.
1SG.DIR book-OBL.F read-PRS-1SG
'I read the book.'

Finally, a closer examination casts doubt on Larson and Samiian's basic claim about the absence of resumptive elements in Kurmanji CPs and

associating it with the relative pronoun status of *ku*. Larson and Samiian first observe that Kurmanji does not allow resumptive pronouns in direct object position in RCs, as shown in (24). Based on the assumption that resumptive pronouns cannot occur in RCs introduced by a relative pronoun, they suggest that the impossibility of resumptive pronouns in this language can only be explained if *ku* 'that' is considered as a relative pronoun.

(24) keçik-a [ku min (*wê) doh dît]
girl-EZ.F that 1SG.OBL (her) yesterday see.PST.3SG
zehf rind bû.
very pretty was
'The girl whom I saw (*her) yesterday was very beautiful.' (Larson and Samiian 2020: 208)

However, in cases of relativization of P-objects, Kurmanji permits the use of Ps that have a pronominal element contracted onto them. Thus, in (25) the form *jê* is an amalgam of the preposition *ji* and the 3SG Oblique pronoun *wê/wî*.

(25) keçik-a [ku min jê ra gul
girl-EZ.F that 1SG.OBL P.3SG.OBL Prt rose
şand] çû Stenbol-ê.
send.PST.3SG go.PST.3SG Istanbul-OBL
'The girl whom I sent roses [to her] went to Istanbul.' (Larsonand Samiian 2020: 208)

If we consider the contracted prepositions as resumptive forms, then a sentence like (25) poses a problem for Larson and Samiian's basic claim about the absence of resumption in CPs in Kurmanji. Zazaki has a distribution similar to Kurmanji in these contexts: again, resumption is not possible with direct objects but available in the prepositional context.[6]

Let us now return to the distribution of EZ in the context of NCCs in Persian. We noted in (19) that EZ is not used in such contexts. However, under the right circumstances, Persian allows for the possibility, largely overlooked in the literature, of using the same particle *-i* used with restrictive RCs in the context of NCCs. The particle was analyzed as an allomorph of EZ

[6] The empirical generalization about the distribution of resumption may need to make reference to the status of the pronominal proforms as *clitics*: those Iranian languages that employ clitics (Persian, Central Kurdish) permit resumption while those that lack clitics (Kurmanji, Zazaki) disallow resumption (except with P-objects, where resumption is forced due to a general ban on P-stranding). We leave a more thorough investigation of the involvement of clitichood in resumption for future research.

in section 3. We see in (26)–(27) that while absence of EZ is grammatical, the use of the *-i* allomorph of EZ is also allowed (cf. (19)).[7]

(26) {*in/un* *edeā /* *(in/un)* *edeā-**yi***} *ke* *vāksan*
this/that claim/ this/that claim-**i** that vaccine
xatarnāk=e=ro *man* *matrah* *na-kard-am.*
dangerous=COP.PRS.3SG=RA 1SG mention NEG-did-1SG
'I didn't bring up the claim that the vaccine is dangerous.'

(27) {*in/un* *shāye'e/(in/un)* *shāye'e-**yi***} *ke* *vāksan*
this/that rumour/this/that rumour-**i** that vaccine
xatarnāk=e=ro *az* *ki* *shenid-i*
dangerous=COP.PRS.3SG=RA from who heard-2SG
'Who did you hear the rumour that the vaccine is dangerous from?'

The above facts present a further challenge for the case analysis of EZ, as the mere claim that Kurmanji CPs are [+N] and Persian CPs are [–N], itself already questionable, will not make the right prediction about the distribution of EZ in the NCC context. The case analysis would need to assume that CPs can be optionally [+N] or [–N] in Persian. There is no morphosyntactic distinction (or any other independently attested difference) between the CPs following the head N in the contexts with or without EZ to support this claim.

In the syntactic literature on NCCs, two main camps can be identified with respect to the analysis of these structures. One camp (Stowell 1981; Napoli 1989: 250; den Dikken 2006: 244) takes the NCC to serve at the subject of predication for the projection of the head noun (cf. '[that S] is the claim'), with the surface order derived as a result of inversion of NP around CP, as illustrated in (28).

(28) [$_{DP}$ *the* [$_{FP}$ [$_{NP}$ *claim*] [F [$_{RP}$ [$_{CP}$ *that S*] [RELATOR t_{NP}]]]]]

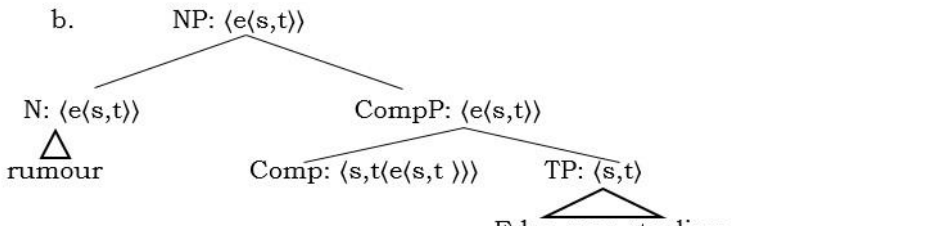
According to another prominent approach (Kratzer 2006; Moulton 2009, 2013; Krapova and Cinque 2015), the NCC serves as (a subpart of) the predicate for the projection of the head noun. Under this view, the surface order matches the base-generated order of constituents and no inversion is

[7] An anonymous reviewer points out that there are speakers who find examples (26) and (27) unnatural with *-i*, even in the discourse contexts we specify later in the text. We have checked these facts with several native speakers and they found the examples with *-i* acceptable so long as the clause is discourse-anaphoric or hearer-old. As shown in (26)-(27), the use of a demonstrative is obligatory in the examples without the particle *-i*, but optional in those with *-i*. Given this contrast, the examples with *-i* may be more readily accessible without the use of a demonstrative, as the absence of a demonstrative forces the use of *-i*.

involved. The *CP Predicate Hypothesis* (Kratzer 2006; Moulton 2009) has the complementizer turn the clause into a predicate: (29).[8] The Comp identifies the content of the noun with the proposition it embeds; e.g., the content of the rumour is the proposition that Edna was stealing, and the CP combines with the noun by predicate modification.

(29) a. rumour that Edna was stealing

b.



(Moulton 2013: 277)

On the assumption that (29) is correct as given, there is no functional head present between the noun and the CP and there is also no inversion. As we have discussed above, we see EZ as the exponent of the inversion process in syntax. The strategy in (29) is thus expected to give rise to absence of EZ. This matches Persian (19) and the versions of (26) and (27) that lack *-i*. These examples are outputs of (29). The versions of (26) and (27) that DO contain *-i* then likely differ in their syntax from their 'bare', EZ-less counterparts. The inversion operation that manoeuvres the NP around the CP in (28) is responsible for the emergence of EZ, in line with Kahnemuyipour (2014). Thus, the versions of (26) and (27) with EZ are outputs of (28).[9]

Importantly, the versions of Persian (26) and (27) with EZ differ from their 'bare' EZ-less counterparts not only in their syntax but also in their interpretation: the NCC in the versions of (26) and (27) with EZ is interpreted

[8] For Krapova and Cinque (2015), the predicativity of the NCC is a function of relativization: the NCC is treated as a subpart of a relative clause with a silent copula and a null relative pronoun (*the claim WHICH IS that S*). Though the Persian and Kurmanji NCC data are compatible with this analysis, we do not follow it in the text because support for the postulation of a relative clause with a silent copula and a silent left periphery is minimal and equivocal.

[9] Our analysis of NCCs in Persian relies on a head-initial syntax, mirroring the structure of verbs taking a complement clause. This could suggest that the kind of flexible headedness proposed for Persian verb phrases (see, for example, Karimi 2005) might extend into the nominal domain. We are grateful to a reviewer for bringing this to our attention.

as hearer-old, discourse-anaphoric.[10] This falls out naturally from a derivation along the lines of (28). The NCC is base-generated as a subject of predication. In syntactic situations in which a particular constituent can in principle be structurally represented either as a subject or as a non-subject (cf. the active/passive alternation), construal of this constituent as a subject shows a strong tendency to deliver a topical, hearer-old interpretation.

Turning to Kurmanji and Zazaki, one can conclude that only the strategy in (28) is used for the formation of NCCs, and as a result, EZ is always required. We noted at the outset that Kurmanji and Zazaki EZ is sensitive to the phi-features of the head noun, which we take to be a case of agreement. As such, one may relate the obligatory use of the inversion strategy in (28) (and the presence of EZ) to the requirement in Kurmanji and Zazaki morphosyntax to engage in a phi-feature agreement relationship with the head noun, exponed on EZ. This process can only be an outcome of (28) (as opposed to (29)). The Kurmanji and Zazaki examples in (17) and (18), respectively, are thus based on (28). Kurmanji and Zazaki shows no alternation between (28) and (29): the fact that the head noun must engage in phi-feature agreement with EZ entails that Kurmanji and Zazaki NCCs can only avail themselves of (28), in which the NCC is a subject. Because in both languages the NCC has no choice but to be syntactically represented as a subject (and consequently there is no alternation in this language between (28) and (29)), there is no information-structural effect associated in Kurmanji and Zazaki with the use of (28). As a result, the NCC in (17), (18) and similar such constructions in both languages can be either hearer-old or hearer-new.

Indeed, the correlation established here between NCC syntax and the presence of agreeing EZ finds further support from Central Kurdish (CK). For

[10] The EZ-less versions of (26) and (27), by contrast, are usable in both hearer-old and hearer-new contexts. This interpretive contrast between 'bare' and morphologically more complex NCC constructions is similar (though not identical) to the one Hankamer and Mikkelsen (2021) discuss with reference to the two types of NCC constructions found in Danish. In both Danish and Persian, the morphologically more complex version (employing EZ in Persian and a preposition in Danish) is only compatible with a construal of the information conveyed by the NCC as hearer-old. Danish differs from Persian, however, in that its 'bare' NCC construction apparently requires a hearer-new interpretation for the CP. Moulton's (29), from which we have derived EZ-less NCC constructions, is information-structurally neutral. The discursive versatility of Persian EZ-less (26) and (27) is directly in line with this. We will not address here the question of why Danish 'bare' NCC constructions are apparently not as flexible in discourse as their Persian counterparts.

instance, in the Silemani dialect of CK, NCCs are always linked to the head N with EZ (N-EZ CP), as illustrated in (30), similar to Kurmanji and Zazaki.[11]

(30) a. *aw* *bochūn-a* *bāw-á-yī* *ka* *sag*
DIST opinion-EZ common-DEF-EZ that dog
bawafa=ya
loyal=COP.3SG
'the common opinion that dogs are loyal' (Jambrović and Hassan 2023: 16)

b. *aw* *hiwa-ya-ī* *ka* *Shah* *Iran* *je*
DIST hope-DEF-EZ that Shah Iran P
da-hel-et *ba* *pey* *kat*
IND-leave.PRS-3SG P after ever
na-ma
NEG-remain/stay.PRS.3SG
'The hope that Shah would leave Iran faded over time.'

5. Conclusion

We have argued in this paper that the distribution of EZ in the context of adnominal clauses in Persian, Kurmanji and Zazaki follows from the general behaviour of EZ and the syntax of N–CP structures. In doing so, we have shown that the distribution of EZ in the N–CP context in Persian, Kurmanji and Zazaki poses a serious challenge to the case analysis of EZ, which predicts that [–N] modifiers should not require the presence of EZ. The facts from these two languages are instead compatible with the inversion analysis of EZ with the correct understanding of the syntax of N–CP structures. We

[11] Note that restrictive RCs in the Silemani dialect are also preceded by EZ (i), as we see in the other Iranian languages (Persian, Kurmanji and Zazaki) discussed so far. In contrast, non-restrictive RCs are not preceded by EZ in this dialect, and they follow a prosodic break as indicated by the comma (ii):

(i) sag-a bichūk-ak-ān-ī ka á-war-in
dog-EZ small-DEF-PL-EZ COMP PROG-bark-3PL
'the small dogs that are barking' (Jambrović and Hassan 2023: 16)

(ii) sag-a bichūk-ak-ān, ka hamū á-war-in
dog-EZ small-DEF-PL COMP all PROG-bark-3PL
'the small dogs, which are all barking' (Jambrović and Hassan 2023: 16)

There are further complications with respect to the distribution of Ezafe in the context relative clauses in Central Kurdish, e.g. the presence/absence of a complementizer and a possible complementarity with the indefinite marker, which warrant further investigation (MacKenzie 1961, McCaurus 2009, Haig 2019).

have posited that all these languages make use of EZ in the context of restrictive RCs, as expected. In Kurmanji and Zazaki, the regular form of EZ is used, while in Persian, an allomorph of EZ, which appears in the context of CPs, is used instead. With non-restrictive RCs, while Persian does not use EZ, Zazaki uses the regular form of EZ and Kurmanji uses a different type of EZ, known as anaphoric EZ (AEZ). We followed de Vries (2006) in analyzing non-restrictive RCs as restrictive RCs to a silent-headed NP. The distribution of EZ in Persian, Kurmanji and Zazaki non-restrictive RCs follows straightforwardly, as it matches the distribution of EZ following a silent N more generally.

With NCCs, Kurmanji and Zazaki use the regular EZ consistently, while Persian exhibits two options: the allomorph of EZ used with CPs or no EZ at all. We have posited two possible structures for NCCs: (i) NCC as the subject of predication for the projection of the head noun, with the surface order derived as a result of inversion of NP around CP, (ii) NCC as (a subpart of) the predicate for the projection of the head noun, with no inversion involved. While Persian was taken to allow both strategies, Kurmanji and Zazaki allow the former only. The distribution of EZ follows accordingly, with EZ only appearing in structures that involve inversion.

Like Kurmanji and Zazaki (and unlike Persian), several other Iranian languages show some form of agreement with the head N on EZ. In future work, we intend to investigate the distribution of EZ in the context of NCCs in Central Kurdish dialects as well as some other Iranian languages such as Hawrami to assess the tentative connection made here between NCC syntax and the presence of agreeing EZ in a language. More generally, the distribution of EZ in the context of both RCs and NCCs in other Iranian languages needs further investigation to test the proposals advanced in this paper and to gain a better understanding of the syntax of RCs, NCCs and the Ezafe constructions in Iranian languages and beyond.

BIBLIOGRAPHY

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,* vol. 36, pp. 315-332.

Cinque, G. (2010), *The Syntax of Adjectives: A Comparative Study*, LI Monograph 57, Cambridge: MIT Press.

Dikken, M. den (2006), *Relators and Linkers*, Cambridge: MIT Press.

Ghomeshi, J. (1997), "Non-projecting nouns and the ezafe: construction in Persian", *Natural Language & Linguistic Theory,* vol 15(4), pp. 729-788.

Gündoğdu, S. (2023). "Ezafe in Northern Kurdish", *Toronto Working Papers in Linguistics*, vol. 45(1), pp. 1-16.

Gündoğdu, S., & Amine, B. (2023). "Ezafe in Zazaki", *Toronto Working Papers in Linguistics*, vol. 45(1), pp. 1-17.

Haig, G. (2011), "Linker, relativizer, nominalizer, tense-particle: On the Ezafe in West Irania", *Nominalization in Asian Languages: Diachronic and Typological Perspectives*, pp. 363-390.

Haig, G. (2019), "The Iranian languages of northern Iraq", in: *The Languages and Linguistics of Western Asia: An Areal Perspective,* Vol. 6, eds. Haig, Geoffrey & Geoffrey Khan, Berlin: Mouton de Gruyter, pp. 267-304.

Hankamer, J., and Mikkelsen, L. (2021), "CP complements to D", Linguistic Inquiry, vol. 52(3), pp. 1-45.

Jambrović, S., and Mohammed, K. H. (2023), "Nominal linkers in Central Kurdish (Silemānī variety)", *Toronto Working Papers in Linguistics*, vol. 45(1), pp. 1-18.

Kahnemuyipour, A. (2014), "Revisiting the Persian Ezafe construction: A roll-up movement analysis", *Lingua*, vol. 150. pp. 1–24.

Kahnemuyipour, A., and Taghipour, S. (2024), "Ezafe in the context of PPs", oral presentation at the Canadian Linguistic Association Annual Meeting, Carleton University, Canada.

Karimi, S. (2005), *A minimalist approach to scrambling: Evidence from Persian*, Berlin: Mouton de Gruyter.

Karimi, S., and Brame, M. (2012), "A generalization concerning the Ezafe construction in Persian", *Linguistic Analysis,* vol. 38(1), pp. 111-144.

Krapova, I., and Guglielmo, C. (2015), "On noun clausal "complements" and their non-unitary nature", *Annali di Ca' Foscari. Serie occidentale*, vol. 50., pp. 77-107

Kratzer, A. (2006), Decomposing attitude verbs, Ms, University of Massachuetts, Amherst.

Larson, R., and Hiroko Y. (2008), "Ezafe and the Deep Position of Nominal Modifiers", in: *Adjectives and Adverbs: Syntax, Semantics and Discourse,*

eds. McNally, Louise & Christopher Kennedy, Oxford: Oxford University Press, pp. 43-70.

Larson, R., and Samiian, V. (2020), "The Ezafe construction revisited", in: *Advances in Iranian Linguistics*, eds. Larson, Richard, Sedigheh Moradi & Vida Samiian, Amsterdam: John Benjamins, pp. 173-236.

Larson, R., and Samiian V. (2021), "Ezafe, PP and the nature of nominalization", *Natural Language & Linguistic Theory*, vol. 39(1), pp. 157-213.

MacKenzie, D. N. (1961), *Kurdish dialect studies I*, Oxford: Oxford University Press.

McCarus, E. N. (2009), "Kurdish", in: *The Iranian Languages*, ed. Windfuhr Gernot, New York: Routledge, pp. 587–633.

Moulton, K. (2009), *Natural selection and the syntax of clausal complementation*, Amherst, MA: University of Massachusetts dissertation.

Moulton, K. (2013), "Not moving clauses: Connectivity in clausal arguments", *Syntax,* vol. 16(3), pp. 250-291.

Moyne, J., and Carden, G. (1974), "Subject reduplication in Persian", *Linguistic Inquiry,* vol. *5*(2), pp. 205-249.

Napoli, D. J. (1989), *Predication theory: a case study for indexing theory*, Cambridge: Cambridge University Press.

Samiian, V. (1994), "The Ezafe construction: Some implications for the theory of X-bar syntax", *Persian Studies in North America*, pp. 17-41.

Samiian, V., and Larson, R. (2023), "Middle Persian ezafe", in: *Advances in Iranian Linguistics*, vol.2, eds. Simin Karimi, Narges Nematollahi, Roya Kabiri and Jian Gang Ngui, Amsterdam: John Benjamins, pp. 100-129.

Samvelian, P. (2007), "A (phrasal) affix analysis of the Persian Ezafe", *Journal of Linguistics,* vol. 43(3), pp. 605-645.

Shlonsky, U. (2004), "The Form of Semitic Noun Phrases", Lingua, vol. 114(12), pp. 1465-1526.

Shlonsky, U. (2010), "The Cartographic Enterprise in Syntax", *Language and Linguistics Compass,* vol. 4(6), pp. 417-429.

Skjærvø, P. O. (2009), "Middle West Iranian", in: *The Iranian Languages*, ed. Windfuhr Gernot, New York: Routledge, pp. 196-278.

Stowell, T. (1981), *Origins of phrase structure*, Cambridge, MA: MIT dissertation.

Taghipour, S., and Kahnemuyipour, A. (2023), "Nominal Linkers in Iranian Languages"*, Toronto Working Papers in Linguistics,* vol. 45(1).

Vries, M. de (2006), "The syntax of appositive relativization: On specifying coordination, false free relatives, and promotion", *Linguistic Inquiry*, vol. 37(2), pp. 229-270.

YEREVAN STATE UNIVERSITY PUBLISHING HOUSE

JIL 1 (2024): 142-162

Journal of Iranian Linguistics

Mirativity in Persian

Mohammad Rasekh-Mahand
Bu-Ali Sina University

doi.org/ 10.46991/jil/2024.01.07

Abstract: Mirativity, as a distinct grammatical category, can be marked by different markers and strategies. In this paper, it is argued that, contrary to previous studies, Persian marks mirativity by using morphosyntactic forms. Three different grammatical tools are identified. First, it has a sentence final clitic '=*ā*' used as mirative marker on its own right. It indicates that the information is newsworthy, unexpected and surprising. Second, the sentence final particle '*ke*', among its different functions, marks mirativity, as well. Third, using different perfect verb forms in Persian is a mirative strategy, which is strongly connected to indirect evidentiality. The data from Persian widens our understanding of mirativity cross-linguistically, showing that a language can have different ways to mark it simultaneously.

Keywords: mirativity, evidentiality, perfect verb, focus, grammaticalization

Mohammad Rasekh-Mahand
E-mail: rasekh@basu.ac.ir

Received: 16.01.2024
Revised: 20.06.2024
Accepted: 14.07.2024



This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License

1 Introduction

Mirativity, in typological studies, first appeared as a by-product of the studies on evidentiality and was defined as a category whose function is to report information which is new or surprising to the speaker (DeLancey 1997). DeLancey (1997) introduced mirativity as a new descriptive category distinct from evidentiality, providing examples from different languages. He argued

that his overview can contribute to more widespread documentation of mirativity in different languages. After two decades, typologists spotted many mirative markers cross-linguistically and widened our understanding of this category (e.g., Lazard 1999, Aikhenvald 2012, Delancey 2001, 2012, Hengeveld and Olbertz 2012, Fang 2018). They mostly tried to demonstrate how mirativity is different from evidentiality and other grammatical categories. Aikhenvald (2012) has probably conducted the broadest typological study of mirativity up to now. She emphasizes the independence of mirativity from evidentiality and other categories and argues that the in-depth studies of mirative marking in different languages show that the category embraces the following values: (i) sudden discovery, sudden revelation or realization, (ii) surprise, (iii) unprepared mind, (iv) counter-expectation and (v) new information. All these values can refer to (a) the speaker, (b) the addressee, or (c) the main character [of the story] (Aikhenvald 2012: 437). These different mirative meanings can be expressed formally by (a) a complex verbal construction, (b) a special verbal affix or a particle, and (c) a special series of pronouns and other means (Aikhenvald 2012: 438). Aikhenvald (2012) provides data from a variety of languages, showing different formal ways of expressing different values of mirativity. As an example, in Kham (Sino-Tibetan) a complex verbal construction marks mirativity (ex. 1). The context for this example is that the speaker had invited guests to his house, and Jhupurya also shows up uninvited or unanticipated. The host has uttered this sentence using a complex verbal construction to announce his arrival. The verbal suffix, *-wo*, marks the mirativity in this sentence.

(1) *Juhpurya* ***u-hu:u-wo***
Jhupurya 3SG-come-PFV.NMLZ

o-le-o Kham (Aikhenvald 2012: 442)
3SG-be-NMLZ
'Jhupurya has arrived!'

Mirativity does not specify any information source, and it can be used with direct evidence, inference, etc. Aikhenvald (2012: 475) argues that mirativity, recognized as a separate concept by DeLancey (1997), is a valid notion, which allowed typologists and grammarians to study it cross-linguistically and identify different meanings and strategies for it.

While mirative meanings can be expressed by lexical means in any language, the number of languages that have grammaticalized it is much fewer (Aikhenvald 2012). Among the second group, some of them use a distinct marker to indicate it, such as a verbal affix, a particle, etc., but other languages have 'mirative strategies' (Aikhenvald 2012: 436), i.e., grammatical

markers whose main function is to show other categories but they express mirative meanings in certain contexts.

There have been some studies on evidentiality in Persian (Lazard 1999, 2001, Jahani 2000, Utas 2000), however, the literature on mirativity is not widespread (see section 2) and no grammatical means to mark mirativity is reported. Lazard (1999) is the only exception who refers to mirativity, but argues that this category is not grammaticalized in Persian. Studying mirativity in some South-Eastern Europe and Western Asia languages, he argues that Persian perfect, while showing evidentiality, is not a good candidate for marking mirativity.

The aim of this paper is to show that Persian has both a mirative marker and mirative strategies. Aikhenvald (2012:458) observes that it is possible for a language to have several forms which express different values of mirativity. Accordingly, I will show that Persian marks mirativity in more than one way. First, the sentence final clitic '*=ā'*, whose function is neglected in previous studies, acts as a mirative marker in this language in many contexts. Moreover, the particle *'ke'*, when appearing sentence-finally, can mark mirativity. In addition, the perfect verbal form marks mirativity, among its other functions. Identifying and documenting these markers and strategies will widen our understanding of mirativity cross-linguistically.

The organization of the paper is as follows. Section 2 reviews the related literature on Persian, especially on evidentiality. In section 3, I argue that a sentence final clitic, *=ā, (*also pronounced *=yā,* or *=hā),* is a bifunctional morpheme, which encodes mirativity, as one of its functions in Persian. In section 4, I show that the particle '*ke'*, among its different functions, is a mirative marker when it appears at the end of sentence. Section 5 is about perfect verbal form in Persian and its use as a mirative strategy. I show that this form, in addition to marking indirect evidentiality (contrary to Lazard 1999), marks new and unexpected information which causes surprise. Section 6 is the conclusion.

2 Previous literature on Persian

Linguists have studied evidntiality in Persian (e.g., Jahani 2000, Utas 2000) and other Iranian and neighboring languages (see Comrie 2000 for on an overview and papers in Johanson & Utas 2000). However, since mirativity is a new concept in linguistic studies, it is not discussed widely in Persian (Lazard 1999).

Lazard (1999) believes that in languages of South-Eastern Europe and Western Asia three values of hearsay, inference and unexpected observation fall within the cover category of mediative. For him mediatives "only interpose an unspecified reference to the origin of the information between speaker and his discourse" (Lazard 1999:96). However, he emphasizes that "no definite example of the mirative has been reported in Persian' (Lazard 1999: 99), and the mediatives always refer to past and they can be interpreted as resultatives and inferential, like example (3):

(2) *bārān qat šod-e ast*
rain cut become-PTCP be.3SG
'The rain has stopped.'

He believes that in example (2) the meaning associated with perfect form is not mirative, but evidential. He reasons that since mediative forms always refer to the past in Persian, they could be equally interpreted as resultative or inferential.

He finds this term mediative more appropriate than evidential or mirative; however, it has not been widely adopted. Regarding Persian, he observes that the mediative form, (in this case perfect verbs), "include not only hearsay and inference, but also the experiential (i.e., a retrospective view of past events, as distinct from the resultative) and what I have called the remote past or completed past, even in the 1st person" (Lazard 1999:99). He provides example (2):

(3) *man ālmāni harf mi-zad-e-am*
I German word DUR-beat-PTCP-be.1SG

ammā hālā farāmuš kard-e-am
but now forgetting do-PTCP-be.1SG
'I used to speak German, but now I have forgotten it'.

In example (2), the perfect verbs are not marking hearsay or inference, since the sentence refers to the first person. Hence, Lazard calls its function remote past. So, generally Lazard (1999) argues that (a) there is no specific mirative marker in Persian, and (b) the perfect verb form in this language does not mark mirativity. He argues that the use of perfect verb forms as evidential markers does not indicate the specific source of information, but is just in opposition to sentences that indicate nothing about the source of information (Lazard 2001: 362). He provides the following example to support his argument:

(4) *sob-e* *sahar* *Nane* *dide-bud=eš,*
morning-EZ dawn PN had.seen-3SG

bāzam *jelo* *xune* *rāh* *mirafte*
again before house way was.going
'At dawn Naneh had seen him, he was again walking in front of the house.'

Lazard argues that the verb form mirafte functions as an evidential, indicating that the information was originally obtained from a source named Naneh. This information is not new to the speaker and he is not surprised by it or doubtful about its accuracy. The speaker is simply reporting it as hearsay. However, the other verb form dide bud, which likely also reflects Naneh's original words, is not an evidential. This means that the information conveyed by *dide bud* is considered equally old or new as the information conveyed by mirafte, but the speaker does not feel the need to mark it as hearsay because it is not significant. In the next sections, I will argue that both of Lazard's findings regarding Persian can be challenged.

Evidentiality in Persian is discussed (though under different names) by several scholars. Windfuhr (1987) is among the earliest scholars who states that some of the verb forms which refer to remote past in the literary register are used in colloquial language to express the category of inference, that is mainly second-hand knowledge, conclusion and reminiscence. Jahani (2000) argues similarly that perfect form of the verb is preferred form for the inferred and reported information, but for eye-witnessed information, both perfect and simple past are used. She concludes that perfect form is not fully grammaticalized, and among its other functions, it can indicate indirect evidence. Utas (2000) who calls the utterances which report non-witnessed action 'epistemic', admits that certain perfect form of verbs in Persian show this epistemic information, while they have other functions like resultative or aorist. However, he argues that in some of the derived forms, the epistemic component is dominant; such as "past perfect" (ex.5), and a "durative perfect" (ex. 6):

(5) *kard-e* *bud-e-* *ast.*
do.PPART be.- PPTCP AUX
'He had done.'

(6) *mi-kard-e* *ast*
IMP-do-PPTCP AUX
'He has had been doing.'

In the following example, the perfect verb form '*raft-e-ast*' (has gone) shows that the speaker has heard the news, not directly observed:

(7) *šenid-e-am* *ke* *ahmad* *diruz*
hear-PTCP-1SG that Ahmad yesterday

safar *raft-e-ast* (Utas 2000:232)
journey go-PTCP-be.3SG
'I've heard that Ahmad has gone on a trip yesterday.'

However, since the matrix verb is an evidential verb, which means "I have heard", it is not easy to argue that the perfect form of the embedded verb and not the matrix verb gives rise to the epistemic meaning.

With regard to mirativity, Perry (2000:236), in line with Lazard (1999), argues that among different varieties of Persian, the perfect is only used to mark this category in Tajiki Persian (spoken in Tajikistan). They clearly state that (Iranian) Persian perfect does not have mirative meaning (see section 5).

In sum, the scholars studying Persian morphosyntax have not identified any marker or strategy of mirativity. They generally believe that one of the functions of different perfect forms of the verbs is to mark hearsay or inferential evidence. In this paper, this proposal is challenged.

3 Sentence final clitic '=*ā*'

DeLancey (1997:49), on mirativity, states that "languages differ not in whether they have means to express it, but in the degree to which its expression is integrated into the grammar". One of the mirative markers which occurs in a number of languages is verbal affixes or particles (Aikhenvald 2012:446). Persian has a sentence-final clitic which in colloquial speech and takes the form *=ā* (*=hā/=yā* in postvocalic contexts). This form has remained understudied and unanalyzed. Since Persian is a verb-final language, in most cases *=ā* attaches to the verb, but in sentences which are not verb-final, it attaches to the last element of the sentence. This morpheme has more than one function and I argue that one of its functions is to act as a mirative marker. I will show that it is an emphatic marker, too, and there is another function for *=ā* to mark vocatives, e.g., *vālā hazrat=ā* (her majesty!) (Lazard 1957:103). It also used to be added to different words to mark sympathy or as an honorific marker in old texts of New Persian, but it is not used in this way anymore.

The following examples show that the sentence final *=ā* is a mirative marker and mirativity is integrated into the Persian grammar system. In the examples in (8), which involve change of state verbs, the meaning associated with *=ā* is newsworthiness and surprise.

(8) a. *barq qat' šod=ā*
power cut become.PST.3SG-MIR
'The power went off'

b. *belaxare qabul kard=ā*
finally accept do.PST.3SG-MIR
'Finally, he/she accepted (it).'

c. *bozorg šod-e=hā*
old become-PCPT-MIR
'He is grown up.'

In (8a), the speaker informs the addressee that some change of state happened. The presence of *=ā* here shows newsworthiness and surprise in being an unexpected situation. In (8b), the speaker did not expect the person referred to by the subject to accept (it), and now reports this as news, accompanied with surprise. And in (8c), the speaker shows surprise regarding the person referred to by the subject. All of the examples in (8) show a new state which is newsworthy and surprising to the speaker and/or the addressee.

In (9), no change of state is observed and the sentences simply report facts.

(9) a. *šenid-am āb*
hear.PST-1SG water

sard-e=hā, sarmā na-xor-i
cold-be.PRS.3SG=MIR cold NEG-eat.PRS-2SG

'I have heard the water is cold, be careful not to get cold."

b. *hendune širin-e=hā*
melon sweet-be.3SG-MIR

'The melon is sweet.'

In (9a), the speaker shows surprise, stating that the water (for shower) is unexpectedly cold and asks the addressee to be careful and not get cold. In (9b), the speaker is simply stating a fact about 'the melon' with surprise (and maybe inviting the addressee to eat it). In all of the examples in (8) and (9),

=ā indicates that the sentence carries a new information, with an overtone of surprise, whether the verb is a change of state verb or simply reporting a fact. So, *=ā* here marks mirativity. If it does not appear in these sentences, they lose the mirative meaning, but they remain grammatical. No other subtle meaning of tense, aspect, or modality is detected with its occurrence.

While mirativity and evidentiality are often connected cross-linguistically, these two categories are not universally expressed by one and the same morpheme (de Hann 2012). The Persian morpheme, *=ā,* does not inform the speaker on the evidence of the news in sentence and is not a marker of evidentiality. For example, while in (9b) the evidence is directly presented (the speaker is eating the melon), in (9a) the speaker has heard that the water is cold and not directly observed.

Peterson (2017) presents a test, to show if a form entails mirativity or not. It employs negation, and if a form like *=ā* marks mirativity, then it is not part of the propositional content of the sentence and thus the negation of the predicate will not affect the mirative meaning of the sentence. For example, the negation of (9b), presented in (10), shows clearly that the mirative meaning is retained, while the propositional meaning is reversed.

(10) *hendune* *širin* *nist=ā*
melon sweet NEG.be.3SG-MIR
'The melon is not sweet.'

In Persian, the mirative marker =ā can be accompanied with exclamative particle *vāy*, as in (11). De Haan (2012) regards this as a feature which shows that the sentence shows mirativity.

(11) *vāy* *sard šod-e=hā*
EXC. cold become-3SG-MIR
'It became cold.'

Newness and surprise go together. New information has some piece of surprise in itself. Normally, the newness of information is associated with time. Events that have happened in the present or recent past are better candidates of carrying new and unexpected information than those in the remote past. Therefore, mirativity is more frequent in sentences referring to present or recent past times. In (12) the time of happening is the very recent past and the speaker reports that the child ate too much and s/he got sick.

(12) *in hāleš* *xarāb* *šod=ā*
this health ruin become.PST.3SG-MIR
'S/he got sick!'

Marking mirativity is not the sole function of sentence final morpheme '*=ā*' in Persian. The second function of this clitic which needs to be distinguished from marking mirativity is that it adds emphasis to imperative and prohibitive sentences. In the following examples, the form of the sentences is imperative, and *=ā* does not change the propositional meaning, but adds emphasis in doing or not doing the action:

(13) a. *na-r-i=yā*
NEG-go-2SG-EMP
'Don't go.'
b. *qand* *na-xor-i=yā*
sugar NEG-eat-2SG-EMP
'Don't eat sugar.'

In the same line, Amoozade and Tavangar (2009) show that deontically-oriented past-tense forms can be used for the expression of direct orders in Persian. In this construction, '*=ā*' can be used again to show emphasis. Let us consider the following conversational exchange:

(14) a: *be-r-am* *kebrit* *be-xar-am?*
SUBJ-go-1SG matches SUBJ-buy-1SG
'May I go and buy matches?'

b: *raft-i* *umad-i=yā*
go.PST-2SG come.PST-2SG-EMP
'Go and come (soon).'

The function of '*=ā*' in this sentence is to add emphasis and to ask the addressee to do it 'very soon'. So, the function of '*=ā*' in imperative or prohibitive sentences or past-tense forms which express direct orders in Persian is not to show surprise or new information, but to emphasize and urge the addressee to do (soon) or not do the action. I have glossed it as EMP, instead of MIR.

There are some instances of sentences with '*=ā*' in which both of the functions discussed above are observed or at least difficult to separate. As an example, in a context in which a family is waiting for guests and are preparing food for them, they suddenly notice that the guests are very near. The daughter of the family utters example (15):

(15) *Ali* *inā* *resid-an=ā*
Ali others arrive.PST-3Pl-MIR.EMP
'Ali and others are arriving.'

Here, *=ā* can have two functions. The speaker gives the overtone of surprise to the family on early arrival of the guests. On the other hand, she urges the family to do their job fast and prepare the food.

In another occasion, the wife brings the empty bottle of jam and says the following sentence to her husband:

(16) *morabbā* *tamum* *šod=ā*
jam finish become.PST-MIR.EMP
'The jam has finished'.

The sentence has new information in it with surprise, and at the same time, it is an order for buying jam. So, in many cases, two functions of '*=ā'* occur together. This co-occurrence is not strange, since imperatives are orders which will be done in near future and normally, they have new (and sometimes unexpected) information for the addressee. These similar functions are achieved by a single form.

In sum, the sentence final clitic *=ā* in Persian is a mirative marker on its own right, which encodes the information as newsworthy or surprising and frequently refers to current situation. It has another function, namely to emphasize the order or avoidance in imperative sentences. These two functions sometimes occur simultaneously in this marker.

Before ending this section, it is noteworthy to show that *=ā* as a mirative marker is also found in other Iranian and non-Iranian languages of the area, yielding support for contact-induced copying of these neighboring languages. While there have been some studies on evidentiality and related matters in Iranian, Turkic and beyond (Johanson and utas 2000, Haig and Khan 2018), *=ā* or similar forms are not reported as mirative (or evidential) marker in these languages[1]. However, different forms similar to Persian *=ā* (or exactly the same form) are found in some of the Iranian and neighboring non-Iranian languages in Western Asia. For example, in Tati (17), Talyshi (18), Gilaki (19), Mazandarani (20) and Central Kurdish (21) which are spoken in north and north-west of Iran, this final marker is '*=(y)e'* or '*=(h)ā'*:

(17) *a dâr xəšk âbe-ye*
that tree dry become.3SG-MIR
'The tree dried out.'

[1] There are few studies on the languages of Western Asia which refer to mirativity. Van der Wal Anonby (2018:633) asserts that in Kumzari (an Iranian language spoken in Oman), one of the verb forms is mirative, which lacks any formal marker rewrite as this is unclear. Anonby and Taheri-Ardali (2018:757) report that in Bakhtiari (an Iranian language), the non-past can also be used with a mirative extension.

(18) *Samad əšta bâɣi furutəša=ye.*
Samad his garden sell.PST.3SG-MIR
'Samad sold his garden.'

(19) *barf bame-ye*
snow came.3SG-MIR
'It has snowed.'

(20) *Ali burd= ā*
Ali go.PST.3SG-MIR
'Ali went!'

(20) *gešt=yān l-ena bu-mn-ā*
all=3PL Direct-Place-DEM be-3rd.PL-MIR
'Everyone was there.'

Among non-Iranian languages in Western Asia, the evidential (and/or mirative) marking is discussed vastly in Turkish varieties (Slobin & Aksu 1982, DeLancey 1997, Johanson 2012, Bulut 2018). They mostly refer to '*-mIš*' perfect, which shows resultative, inferential and mirativity in Turkish. But in Turkish studies literature, there is no mention of a separate mirative marker, like *=ā*, which appears in Persian. Bulut (2018:424) argues that while the Turkish perfect '*-mIš*' is used to mark indirect evidentials and also mirativity (DeLancey 1997), it only marks resultative in Turkic varieties of Iran (see also Kiral 2000 for the same observation in Khalaj). However, 'inferential or evidential connotations are expressed by the evidential/inferential '*-ImIš*', as in '*yatmiš-ImIš*', 'she had obviously gone sleep'". Johanson (1998) observes that the fact that '-mIš' forms do not signal inferentiality in Irano-Turkic varieties seems to be due to Persian influence (for a different view on Azeri Turkish see Lee 1996:49).

However, I found that *-ā* has the same function as it has in Persian, in a variety of Turkish spoken in southern parts of Hamedan in west of Iran. Among the Turkish varieties of Iran, Turkish speakers in south of Hamedan, which is genetically from South Oghuz or Afshār branch of Turkish language group (Bulut 2018), use the *-ā* form to express mirativity. The *-ā* form can be added to these constructions to show surprise, as in Azeri Turkish in (22):

(22) *Ali yāt-mIš-ImIš-ā*
Ali sleep-PRF-COP-MIR
'Ali has been sleeping.'

Another suffix which makes perfect in this area is *-ib* (Bulut 2018:424). This form also can accompany with *-ā* to mark mirativity, as in (23):

(23) *dost-om* *gal-ib-di-yā*
friend-1SG come-PRF-COP.3SG-MIR
'My friend has come.'

The *=ā* as mirative marker is not used only in perfect sentences. It can be used in other tenses, too (e.g. 24):

(24) *yāqor-o-yā*
rain-3SG-MIR
'It is raining.'

The interesting point is that *=ā* is also used in emphatic orders (25) and deontically-oriented past-tense forms which mean can be used as orders (26):

(25) *gal-ā*
come-MIR
'Come!'

(26) *gal-d-i-yā*
come-PST-2SG-MIR
'You came!'

It can be observed that the form *=ā* and its pattern is replicated in the Turkish variety spoken in this region.

Reportedly, the same form *'-(h)ā/-(h)a'* is used in Azeri Turkish, too, for a warning or admonition (27), or expressing surprise (28) (Lee 1996:89) :

(27) *olar-a* *bir söz* *de-mə-ha!*
they-DAT one word say-2S.NEG.IMP-MIR
'Don't say any thing to them, okay?'

(28) *Gözəl-di(r)* *ha!*
pretty-be.3S.PR MIR.
'Isn't it pretty?'

So, the same morpheme with the same function is found in some varieties of Turkish language in Iran. However, as far as I know, it is not reported in other varieties in other areas. It seems that Turkish varieties have replicated this mirative marker from Persian (or other Iranian languages).

4 *'ke'* in sentence final position

The word *ke* has a variety of grammatical functions in Persian. Its main function is to mark subordinate clauses. It functions as a relativizer in relative clauses (29), and as a complementizer in complement clauses (30).

(29) *pesar-i* *ke* *did-i* *mariz* *ast*
boy-RELM KE see.PST-2SG ill be.PRS.3SG
'The boy whom you saw is ill.'

(30) *mi-dān-am* *ke* *Ali* *raft-e-ast*
IND-know.PRS-1SG KE Ali go.PST-PRTC-be.PRS.3SG
'I know that Ali has gone.'

This word is also used as focus marker, appearing after different kinds of constituents in a sentence. In this function, *ke* focalizes the constituent which follows it and makes it prominent (Ghomeshi 2013, Oroji and Rezaei 2013:80).

(31) *man* *ke* *ketab* *ro* *be* *Ali* *ne-mi-da-m.*
I KE book OM to Ali NEG-IND-givePRS-1SG
'I won't give the book to Ali.'

(32) *man* *ketab* *ro* *ke* *be* *Ali* *ne-mi-da-m.*
I book OM KE to Ali NEG-IND-givePRS-1SG
'I won't give THE BOOK to Ali.'

(33) *man* *ketab* *ro* *be* *Ali* *ke* *ne-mi-dœ-m.*
I book OM to Ali KE NEG-IND-givePRS-1SG
'I won't give the book to ALI.'

In addition to these functions, when *ke* occurs sentence-finally, it plays different roles. If it appears after interrogative sentences, it adds some rhetorical nuances. For example, in (34) the speaker is not asking a real question, but he wants an affirmative response from the addressee:

(34) *šām* *xord-i* *ke* ?
supper eat.PAST-2SG KE
'You have had supper, haven't you?' (Clearly expecting a positive answer)

If *ke* appears at the end of declarative sentences, it marks an unexpected situation which surprises the speaker (and addressee). For example, in a context that the participants did not expect Ali to pass the exam, and one of them finds that he did, he utters the following sentence (35):

(35) *Ali pās kard ke!*
Ali pass do.PST.3SG MIR
'Ali passed the exam!'

The addressee did not expect Ali to pass the exam and this news has surprised him. In this sentence, *ke* can be omitted without affecting the grammaticality of the sentence. But, in that case, the sentence turns to a simple news and it loses the effect of additional surprise. If, as a test, we employ negation, the negation of the predicate will not affect the mirative meaning of the sentence and it shows that *ke* acts as a mirative marker in this sentence. The following are some more examples of using *ke* as a mirative marker. (37) is the negative form of (36) in which the mirative meaning is not affected.

(36) *'e barf umad-e ke*
wow snow come.PST-PTCP MIR
'Wow, it has snowed!'

(37) *'e barf na-yumad-e ke*
wow snow NEG-come.PST-PTCP MIR
'Wow, it hasn't snowed!'

(38) *barq qat šod ke*
power cut become.PST.3SG MIR
'The power cut off!'

In (36) and (37), the exclamative marker *e* appears at the beginning of the sentence, emphasizing the unexpectedness of the news, and (38) is an unexpected change of situation.

In this function, *ke* can be replaced with *=hā*, showing that they have the same function when used as mirative marker. In addition, they could not occur in the same sentence, representing another evidence that they have a similar function (35):

(39) **barq qat šod ke= hā*
power cut become.PST.3SG MIR-MIR
'The power cut off!'

With regard to the above discussion, I conclude that one of the functions of *ke* in Persian is to mark mirativity.

5 Perfect form of verbs

While a few languages have special marker for mirativity, languages express mirative meanings through other grammatical categories. DeLancey (1997, 2001) refers to these as “mirative as a semantic space”; and Aikhenvald (2012:463) uses “mirative strategies” to specify them, "that is, extensions of essentially non-mirative categories which acquire mirative meanings within a given context". Evidentials are among the frequently attested mirative strategies cross-linguistically. DeLancey (1997, 2001) argued that evidentials are associated with the mirative range of meanings. Similarly, Aikhenvald (2012:465) believes that “in small evidential systems, with firsthand (or eyewitness) evidential versus non-firsthand (or non-eyewitness evidential) evidential, non-firsthand typically acquires mirative meanings".

As discussed in section (2), scholars agree that different perfect forms of verbs in Persian, among other usages, can signal evidentiality (e.g., Lazard 1999, Jahani 2000, Bubenik and Ziamajidi 2020, Jügel 2020). They believe that Persian has a small (two-term) evidentiality system, first-hand/direct vs. non-first-hand/non-direct. Before discussing its function, it is needed to introduce its different forms. Persian perfects occur in present or past forms. Table (1) represents the paradigm of different perfects forms in Persian:

Table 1. Perfect forms in Persian

Present perfect	Past perfect
nevešt-e-am write.PST-PTCP-be.1SG	*nevešt-e-bud-e-am* write.PST-PTCP-be.PST-PTCP-be.1SG
nevešt-e-i write.PST-PTCP-be.2SG	*nevešt-e-bud-e-i* write.PST-PTCP-be.PST-PTCP-be.2SG
nevešt-e ast write.PST-PTCP be.3SG	*nevešt-e-bud-e ast* write.PST-PTCP-be.PST-PTCP be.3SG
nevešt-e-im write.PST-PTCP-be.1PL	*nevešt-e-bud-e-im* write.PST-PTCP-be.PST-PTCP-be.1PL
nevešt-e-id write.PST-PTCP-be.2PL	*nevešt-e-bud-e-id* write.PST-PTCP-be.PST-PTCP-be.2PL
nevešt-e-and write.PST-PTCP-be.3PL	*nevešt-e-bud-e-and* write.PST-PTCP-be.PST-PTCP-be.3PL

Different tenses above may omit the final BE-auxiliary in the 3SG subjects, as *'šode-ast > šode',* (become.PST-PTCP-be.3SG*), 'gofte bude-ast > gofte bude',* (say.PST-PTCP-be.PST-PTCP-be.3SG*).* The use of 'be' as an auxiliary in this complex construction in Persian is in line with Aikhenvald's (2012: 445) findings that "complex constructions with mirative meanings involve the verb 'be' or a grammaticalized copula (as in Kham and Magar), or the verb 'become', and 'discover' as in Northeast Caucasian languages, and in Tariana, accompanied by a nominalized verb". There is also a durative perfect form which is similar to other perfect forms, but it uses '*mi-'* to mark durativity (Mofidi & Petre 2022), as well, like (*mi-nevešt-e ast*, DUR-write.PST.PTCP be.3SG).

One of the functions of different perfect forms of the verbs is to mark hearsay or inferential evidence, i.e., non-first-hand (indirect) evidence (41). However, when the sentence is simple past, it means the speaker has direct evidence (40):

(40) *diruz dar jādey-e Tehrān tasadof-e bad-i šod*
yesterday in road-EZ Tehran accident-EZ bad-INDF become.PST.3SG
'A bad accident happened in road of Tehran yesterday.'

(41) *diruz dar jādey-e Tehrān tasadof-e bad-i šod-e*
yesterday in road-EZ Tehran accident-EZ bad-INDF become.PST.PTCP.3SG

While many scholars argue that the perfect verb form shows indirect evidentiality, they do not assert that perfect forms can be among the mirative strategies in this language. Lazard (1999) explicitly asserts that this form makes no mirative overtones. I will argue in this section that Persian perfect verb is a mirative strategy and it is a verbal category which acquires "overtones to do with surprise and information unexpected to the speaker" (Aikhenvald 2012: 463).

In the following examples, the evidence is achieved visually, so the function of the perfect verb cannot signal indirect evidence; however, it is used to mark the surprise of the speaker by seeing an unexpected scene or event. The context for (42) is as follows. The speaker sleeps the night before while the sky was clear; he gets up and opens the window and sees that there is a lot of snow in the yard and says:

(42) *'e, barf umad-e*
Wow, snow come.PST-PTCP.3SG
'Wow, it has snowed (lit.).'

The use of present perfect in this sentence cannot be a strategy for indirect evidence marking, since he is seeing the snow; however, it shows that the speaker is surprised by discovering an unexpected scene.
In a similar context, two women meet each other after a while. One of the women has brought her child. Seeing the child, her friend says:

(43) *xodā jun, če qadr bozorg šode*
God dear, how much old become.PST-PTCP.3SG
'My god, s/he has so grown up.'

Again, the speaker is seeing the child, so the perfect is not used for providing indirect evidence, but for showing mirativity. These sentences show that perfect form can be used to show mirativity in Persian and it is not part of showing evidentiality. Mirativity can be an overtone of perfect forms, while evidentiality is absent. Of course, there are some contexts where the perfect verb can indicate both evidential and mirative meanings. For the following sentence, different contexts can force either of these meanings:

(44) *Ali umad-e*
Ali come.PST.PTCP.3SG
'Ali has come.'

The sentence can show inference or hearsay if the speaker has seen Ali's car in the yard or somebody has told him the news but he has not seen Ali himself. In these contexts, the perfect is used to mark indirect evidentiality. However, if the speaker opens the door and sees Ali unexpectedly, the sentence has mirative overtone. In this way, the perfect marker is very similar to often cited form in Turkish, -miš, which is used to do different functions (Slobin & Aksu 1983, DeLancey 1997). It is important to point out that this sentence is appropriate in context of seeing Ali's car in the yard or hearing from somebody else, but if the speaker hears Ali's car approaching, he cannot use this sentence and instead he should use simple past, Ali umad (Ali come.PST.3SG). When the speaker sees Ali's car approaching, it is direct evidence and the perfect could not be used. When he hears it from somebody else or sees the car in the yard, he gets the indirect evidence and the perfect is used to show indirect evidence. Here, the speaker is using auditory sensory experience as part of Ali's arrival and "his consciousness is involved in the process before its actualization" (Slobin & Aksu 19783: 192). So, in the mirative reading of sentence (44), while the speaker is seeing Ali, he can use perfect form to show the unexpectedness of the event, since the actualization of the arrival is done with no prior consciousness. In example (45), the speaker opens the door and sees Ali. Since his arrival is unexpected, he addresses Ali himself by uttering this sentence:

(45) *be-bin* *ki* *umad-e*
IMP-see who come.PST.PTCP.3SG
'See, who has come!'

This example shows that the perfect verb can be directed to the addressee to show the surprise of the speaker, while the evidence is direct. The examples presented in this section shows that, contrary to Lazard (1999), perfect form in Persian can be classified as a mirativity strategy. In some cases, one of the readings can have mirative value in a proper context, while there are some cases where the evidence is direct (visual), hence the perfect form could not mark indirect evidentiality, but it only has mirative overtone. I conclude that perfect form in Persian is both an evidential and mirative strategy.
The perfect forms in Persian can be used in mirative statements for a different person. While DeLancey (1997:50) asserts that using mirative for first person is odd since "information about the rest of the world may be surprising, but information about oneself should not be"; perfect miratives in Persian occur with first person, if speakers find something surprising for themselves:

(46) *man* *az* *in* *qazā* *xord-e-am*
I from this food eat.PST-PTCP-be.1SG
'I have eaten this food.'

In the above context, the speaker is in a new city and the host has brought him a local food. While eating, she finds that it is not new to her and she has already eaten it, unexpectedly.

6 Conclusion

Persian, like any other language, has different lexical ways to express range of mirative meanings. It uses some lexical items, like '*ta'ajjob kardan*' (to be surprised), exclamative clauses, interjections, like '*e*' and '*ajab*', both equal to English 'wow!', and exclamatory intonation. But, in this paper, I argued that mirativity is encoded in Persian grammar, as well. I showed that two sentence-final forms mark mirativity in this language. The clitic '=*hā*' and the particle '*ke*', among different functions, are mirative markers. In addition to these mirative markers, using the perfect form of the verbs in Persian is a mirative strategy. While this form has different functions, among them showing indirect evidentiality, it can show mirative meaning in specific contexts. These findings provide more evidence for the cross-linguistic finding that languages can use more than one grammatical form to mark mirativity.

BIBLIOGRAPHY

Aikhenvald, A. Y. (2012), "The essence of mirativity", *Linguistic Typology*, vol. 16(3), pp. 435-485. https://doi.org/10.1515/lity-2012-0017

Anonby, E. & M. Taheri-Ardali (2018), "Bakhtiari", in: Haig, G. & G. Khan (eds.). *The languages and linguistics of Western Asia; an areal perspective*, Berlin/Boston: De Gruyter Mouton. pp. 445-480.

Bubenik, V. and L. Ziamajidi (2020), "The rise of the analytic perfect aspect in West Iranian languages", in: Drinka, B. (ed.) *Historical Linguistics 2017. Selected papers from the 23rd International Conference on Historical Linguistics, San Antonio, Texas, 31 July – 4 August 2017*. Amsterdam: John Benjamins Publishing Company, pp. 273-292.

Bulut, Ch. (2018), "The Turkic varieties of Iran", in: Haig, G. & G. Khan (eds.). *The languages and linguistics of Western Asia; an areal perspective*, Berlin/Boston: De Gruyter Mouton, pp. 398-444.

Comrie, B. (2000), "Evidentials: semantics and history", in: Johanson L. & Bo Utas (eds.). *Evidentials: Turkic, Iranian and Neighbouring Languages*, Berlin/ New York: Mouton de Gruyter, pp. 1-12.

de Haan, F. (2012), "Evidentiality and Mirativity." *Oxford Handbooks Online*.

Delancey, S. (1997), "Mirativity: The grammatical marking of unexpected information", *Linguistic Typology*, vol. 1(1), pp. 33-52.

Delancey, S. (201), "Still mirative after all these years", *Linguistic Typology*, vol. 16(3), pp. 529-564. https://doi.org/10.1515/lity-2012-0020

Fang, H. (2018), "Mirativity in Mandarin: the sentence-final particle *Le* (了)", *Open Linguistics*, vol. 4, pp. 589-607. https://doi.org/10.1515/opli-2018-0029

Ghomeshi, J. (2013), "The syntax of pragmaticalization", in: *Proceedings of the 2013 annual conference of the Canadian Linguistic Association.* URL: https://cla-acl.ca/pdfs/actes-2013/Ghomeshi-2013.pdf. Retrieved 02 October, 2024.

Haig, G. & G. Khan (eds.) (2018), *The languages and linguistics of Western Asia; an areal perspective*. Berlin/Boston: De Gruyter Mouton.

Haig, G. & M. Rasekh-Mahand (2022), *HamBam: The Hamedan-Bamberg Corpus of Contemporary Spoken Persian*. Bamberg: University of Bamberg. (multicast.aspra.uni-bamberg.de/resources/hambam/) (date accessed).

Hengeveld, K. & H. Olbertz (2012), "Didn't you know? Mirativity does exist!", *Linguistic Typology*, vol. 16 (3), pp. 487-503. https://doi.org/10.1515/lity-2012-0018
https://doi.org/10.1075/cilt.350.13bub
https://doi.org/10.1093/OXFORDHB/9780195381979.013.0036

Jahani, C. (2000), "Expressions of indirectivity in spoken modern Persian", in: Johanson L. & Bo Utas (eds.), *Evidentials: Turkic, Iranian and*

Neighbouring Languages, Berlin/ New York: Mouton de Gruyter, pp. 185-207.

Johanson, L. (1998), "Code-copying in Irano-Turkic", *Language Sciences*, vol. 20(3), pp. 325-337.

Johanson, L. & Bo Utas (eds.) (2000), *Evidentials: Turkic, Iranian and Neighbouring Languages.* Berlin/ New York: Mouton de Gruyter.

Jügel, Th. (2020), "The perfect in Middle and New Iranian languages", in: Crellin, R. & J. Thomas (eds.), *Perfects in Indo-European languages and beyond*, Amsterdam: John Benjamins Publishing Company, pp. 279-310.

Kiral, F. (2000), "Reflections on -miš in Khalaj", in: Johanson L. & Bo Utas (eds.). *Evidentials: Turkic, Iranian and Neighbouring Languages*, Berlin/ New York: Mouton de Gruyter, pp. 89-101.

Lazard, G. (1957), *Grammaire du persan contemporain*. Translated into Persian by Mahasti Bahreini (2018), Tehran: Hermes Publications.

Lazard, G. (1999), "Mirativity, evidentiality, mediativity, or other?", *Linguistic Typology*, vol. 3, pp. 91–110.

Lazard, G. (2001), "On the grammaticalization of evidentiality", *Journal of Pragmatics*, vol. 33, pp. 358– 368.

Lee, S. N. (1996), *A grammar of Iranian Azarbaijani*, Ph.D. dissertation, Linguistics COGS, University of Sussex.

Meshkinfam, M. (2020), *The grammaticalization of indefinite article in Persian; a Construction Grammar approach*, Ph.D. dissertation. Bu-Ali Sina University, Hamedan, Iran.

Mofidi, R. & P. Petré 2022. "Aspectual marking from a typologically uncommon origin: a quantitative account of the development of *hamē(w)* in Middle Persian", *Folia Linguistica*, vol. 56, pp. 97-127. https://doi.org/10.1515/flin-2022-2025

Oroji, M. R. & A. Rezae (2013), "Exploring 'ke' as a Focus Particle in Persian from both Form and Function Points of View", *Australian Journal of Linguistics*, vol. 33:1, pp. 76-84

Perry, J. R. (2000), "Epistemic verb forms of Iran, Afghanistan and Tajikistan", in: Johanson L. & Bo Utas (eds.). *Evidentials: Turkic, Iranian and Neighbouring Languages*, Berlin/ New York: Mouton de Gruyter, pp. 229-258.

Peterson, T. (2017), "Problematizing mirativity", *Review of Cognitive Linguistics*, vol. 15(2), pp. 312-342.

Slobin, D. & A. Aksu (1982), "Tense, aspect, and modality in the use of the Turkish evidential", in: Hopper P. (ed.), *Tense-Aspect: Between Semantics and Pragmatics*, Amsterdam: Benjamins, pp. 185-200.

Tavangar, M. & M. Amouzadeh (2009), "Subjective modality and tense in Persian", *Language Sciences*, vol. 31, pp. 853-873.

Utas, Bo (2000), "Traces of evidentiality in Classical New Persian", in: Johanson L. & Bo Utas (eds.). *Evidentials: Turkic, Iranian and*

Neighbouring Languages. Berlin/ New York: Mouton de Gruyter, pp. 259-274.

Van der Wal Anonby, Ch. (2018), "Kumzari", in: Haig, G. and G. Khan (eds.), *The languages and linguistics of Western Asia; an areal perspective*. Berlin/Boston: De Gruyter Mouton, pp. 625-658.

Windfuhr, G. (1987), "Persian", In Comrie, Bernard. (ed.), *The world's major languages.* London: Routledge, pp. 523-546.

YEREVAN STATE UNIVERSITY PUBLISHING HOUSE

JIL 1 (2024): 163-183

Journal of Iranian Linguistics

Low vowel dissimilation in Mazandarani

Mohsen Mahdavi Mazdeh
University of Arizona

Sarah Nehzati
University of Tehran

doi.org/ 10.46991/jil/2024.01.08

Abstract: Similar patterns of vowel change in loanword adaptation have been documented for several Iranian languages and language varieties including Mazandarani. However, no convincing accounts of the nature of these processes in Mazandarani have been presented in the literature. We argue that for this language, these vowel alternations are best explained as low vowel dissimilation, a process affecting adjacent syllables with low vowels whereby one of the vowels is raised. Low vowel dissimilation is typologically rare, with the overwhelming majority of the cases identified belonging to the Oceanic family. To show that the vowel changes in question are indeed cases of low vowel dissimilation, we invoke evidence from the language's verbal morphophonology where vowel changes show a more regular behavior and then expand the analysis to loanword adaptation. The dialects discussed in the article are those of Amol, Reineh, and Babol. The two vowels that trigger the process in Mazandarani are the low vowels /æ/ and /ɒ/, but only the former can undergo change. We show that unlike almost all other known cases of this phenomenon, it is the second vowel that undergoes raising in the Mazandarani case in many situations, with this being the preferred way in the dialect of Babol. We end the paper with a discussion of why the two low vowels behave differently, suggesting that /ɒ/'s resistance to change is due to the fact that it is a long vowel phonologically, even if not phonetically.

Keywords: low vowel dissimilation; vowel raising; loanword adaptation; verbal morphology; Mazandarani

Mohsen Mahdavi Mazdeh
E-mail: mahdavi@arizona.edu

Sarah Nehzati
E-mail:
sarahhnehzati@gmail.com

Received: 13.03.2024
Revised: 08.07.2024
Accepted: 30.06.2024

This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License

Introduction

1. Introduction

In this article, we analyze a very frequent form of vowel alternation in Mazandarani that has received very little attention in the literature and propose that it can best be described as low vowel dissimilation. We then examine the theoretically significant ways in which the dissimilation process in Mazandarani differs from similar phenomena identified in other languages.

The body of scholarly work done on the synchronic phonology of Mazandarani is relatively small compared to the other languages of the region with similar numbers of speakers. This may be due to the high degree of similarity between the phonological system of Mazandarani and that of modern Persian, which is a product of not only genetic proximity but centuries of close contact. The two languages have almost identical consonant inventories, with the biggest differences lying in the status of /ɢ/, /ʁ/, and /ʒ/ in some of the two languages' varieties (For more on this, see Borjian 2019). They also have similar syllable structures, with Persian generally allowing for more universally marked coda clusters as seen in (originally Arabic) words such as [æɢl] ("wisdom") and [sæbr] ("patience"), which are resolved in Mazandarani with the addition of epenthetic schwas ([æɢəl] and [sæbər] respectively). The stress systems are also very similar at the word level, with stress generally falling on the last syllable in nouns and adjectives but having a tendency towards the initial position in verbs. The vowel systems are more divergent. However, even there, the difference is most visible in how the sounds correspond (in both loanwords and cognates) rather than the shapes of the vowel inventories. This is visible in the prevalent vowel changes that loanwords entering Mazandarani from Persian typically undergo. This article introduces Low Vowel Dissimilation as the process behind many of these changes. This analysis serves two purposes. First, it accounts for what is arguably the most salient phonological phenomenon setting apart the phonological systems of Persian and Mazandarani, which has often been alluded to but never explained. Second, it introduces a new case of the typologically rare phenomenon of Low Vowel Dissimilation, the study of which has mostly been confined to Oceanic languages (see Section 1.3). We demonstrate that even though the environments that trigger Low Vowel Dissimilation in Mazandarani are the same as those observed in the other few languages in which the phenomenon has been studied, Mazandarani follows a different mechanism in its choice of which vowel to raise (with interesting systematic patterns of further variation among its dialects).

This study examines the dialects of Mazandarani spoken in the urban centers Amol and Babol as well as the small town Reineh (sometimes spelled

as “Rineh”) south of Amol, the dialect of which shows differences in its vowel alternations that are interesting from a theoretical point of view. Amol and Babol are both cities of more than 200,000 residents located near the center of the plains on the southern shores of the Caspian. Reineh is located in the cold mountainous region below the Caspian plain, some 85 kilometers south of Amol. As of 2016, fewer than 1000 people lived in Reineh during winters, but the population reaches several thousand during the summers according to locals, with most of these part-time settlers based in Amol and a smaller percentage based in the capital Tehran (which is 115 kilometers southeast of Reineh). In spite of the close contact with Amol, Reineh has its distinct variety of Mazandarani. The phenomenon under investigation, i.e. Low Vowel Dissimilation, occurs in all three dialects as well as other varieties of the language, with differences in details. Both native words and loanwords are considered in this article, but in native words the discussion is mostly limited to verbs, where low vowel dissimilation can be observed as an exceptionless process interacting with morphology.

1.1. Vowel alternations

The raising of an underlying /æ/ to a mid vowel ([ə] in the dialects of Amol and Reineh, [e] in that of Babol) is the most salient process in Mazandarani’s adaptation of loanwords from Persian. A few examples are presented below (our data sources are discussed in the next section).

(1)

	Persian	Maz. (Amol)	Maz. (Babol)	Gloss
a.	ɢæ̲ˈtɒr	ɢə̲ˈtɒr	ɢe̲ˈtɒr	‘train’
b.	xæˈbæ̲r	xæˈbə̲r	xæˈbe̲r	‘news’
c.	ɒˈdæ̲m	ɒˈdə̲m	ɒˈde̲m	‘human’
d.	xælæ̲ˈbɒn	xælə̲ˈbɒn	xæle̲ˈbɒn	‘pilot’
e.	ziˈnæ̲t	ziˈnə̲t	ziˈne̲t	(female first name)

With the exception of a few cursory remarks, this phenomenon has not been discussed in the linguistic literature. Characterizing the phenomenon in the context of a more general phenomenon of vowel raising occurring in several Iranian languages and language varieties, Kord Zafaranlu and Ezatabadi Pour (2018) present a few examples from the Mazandarani dialect of Babol and argue that the raising process only targets stressless syllables. The examples they provide have the same general structure as example (a) above. However, as examples (b), (c), and (e) in the above table suggest and further examples in the following sections demonstrate, this is not the case in

Mazandarani (although stress might be relevant in determining which vowel undergoes raising; see Section 4). For mostly independent reasons, Modarresi Ghavami (2020) rejects Kord Zafaranlu and Ezatabadi Pour's (2018) characterization. Discussing in the same general family-wide context, she attributes the vowel raising process to syllable structure. She does not offer exact criteria for when raising occurs in languages such as Mazandarani, but attempts to limit the environments by arguing that raising is blocked in closed syllables. Her generalization has important exceptions as we shall see in the following sections, but may hold as a statistical tendency or even as an inviolable constraint in some dialects. However, this leaves the more important question of what triggers vowel change unanswered.

Given the increase of Persian influence in recent decades and the speed of changes resulting from this, loanword adaptation processes in Mazandarani are difficult to study. The high degree of variation across words, dialects, generations, and idiolects in how much a Persian word changes when used in Mazandarani means that finding the prevalent patterns is not always straightforward. To overcome this obstacle, we build the foundations of our proposal by investigating vowel change processes in the productive and exceptionless domain of verbal morphophonology and then use our results to explain the data we observe in loanwords. We argue that factoring out a few independent lexical effects, the vowel alternations observed in Mazandarani loanword adaptation can be viewed as low vowel dissimilation, a process preventing the occurrence of two low vowels in adjacent syllables.

1.2. Data

The core of the observations leading to the present analysis comes from the linguistic knowledge of the authors, both of whom are heritage speakers of Mazandarani (one speaking the variety spoken in Reineh with near-native fluency, the other having a working knowledge of the dialect of Babol, and both of them having years of exposure to the dialect of Amol). However, the entire data presented in this article have been verified through elicitation sessions with native speakers of the language. Thus, the interviews (especially as far as the dialects of Babol and Reineh are concerned) may be viewed as merely a complementary (and confirmatory) source of data.

The interviews were conducted in person in Amol, Reineh, and Babol. We interviewed one male and one female consultant from each of the big cities and only one female speaker from Reineh. The ages of the participants ranged from 30 to 62, and all were born and raised in Amol, Babol, and Reineh. Our speaker from Reineh (82 years old) had lived in Amol for 15 years in her

adulthood (after the age of 50), but as the data shows and the authors' own knowledge of the varieties of Mazandarani in the region confirm, her speech did not show any obvious signs of influence from the Amoli variety. For all of the participants but one of the Amoli speakers, Mazandarani was the dominant language at home throughout the speakers' lives. All participants were bilingual in Persian and Mazandarani with no working knowledge of any other language. For the data on the place of raising in disyllabic words with identical vowels (Section 3.2), we consulted a third Amoli speaker as well (born and raised in Amol).

It must be noted that the use of Mazandarani is rapidly declining in urban areas (see Shahidi 2008 for a detailed report of the situation). Persian influence is ubiquitous and it is in fact difficult to find settings in larger urban areas such as Amol and Babol where entire conversations take place in Mazandarani between younger people without code switching or heavy use of long Persian phrases. This situation results in a lot of inter-speaker variation with respect to loanwords and sometimes makes it difficult to tell apart the use of loanwords from instances of code mixing. Thus, some of the loanword data presented in this paper may occur with higher or lower degrees of change in other speakers' speech.

The goal of the interview sessions was only to confirm the status of the vowels of the words presented in this paper and obtain systematic and reliable data regarding the vowel change under question in the three dialects of the language. The words consist only of verbs and a set of loanwords (from Persian, some ultimately from Arabic). In the case of loanwords, a major worry was that simply presenting the words in Persian and asking for the Mazandarani version might affect the authenticity of the participants' responses. To overcome this issue as much as possible, we divided each interview session into smaller parts, leaving direct questions to the last part and limiting it to words for which other methods had failed.

Each interview session had four parts. First, we asked the participant general questions in order to get a general picture of their speech patterns, especially with regard to Persian influence and patterns of vowel alternation. This part of the interview was conducted in Mazandarani. In the second part, we presented them with Persian verbs (mostly within the context of sentences) and asked them to translate them to Mazandarani. In the third part, we presented the participants with small puzzles. For instance, to get the word for "forest", we would ask them about the name of the vast area with many trees which is home to wild animals. These questions were presented in Mazandarani. Finally, in the fourth part, the words for which the puzzle method was not successful (and had not appeared in the speaker's spontaneous speech either) were presented to the speakers one by one in Persian. It is worth noting that, with only one or two marginal cases, the

answers provided by the participants in this last part did not show signs of having been affected by the Persian prompt.

1.3. Low Vowel Dissimilation

Low vowel dissimilation (henceforth, LVD) is a phonological process affecting adjacent syllables with low vowels whereby one of the vowels is raised. In the most common case, the sequence aCa changes to əCa or eCa. For years, the only cases of LVD introduced in the literature belonged to Oceanic languages. Most notably, in two consecutive works, Blust (1996a, 1996b) did an extensive study on LVD in various Oceanic languages, introducing numerous occurrences of it (mostly as a diachronic process, but in some languages as a productive synchronic process) and reducing the historical sources of the cases to at most five independent instances.

The restriction of the cases to Oceanic languages made it difficult to identify the nature of the process in more detail. As Blust (1996b) pointed out, it was not clear immediately whether certain other circumstances that apply in these languages are inherently related to the nature of low vowel assimilation or not. In particular, in all Oceanic languages discussed by Blust, it is always the first of the two vowels that undergoes change. Moreover, a phenomenon of final vowel loss is observed in all the languages discussed, whose relationship with LVD is not clear.

Further studies by Lynch (2003) and Blevins (2009) shed more light on the subject. Lynch identified LVD as a diachronic sound change process in several other Oceanic languages as well and reduced their historical origins to a few cases. He also showed that the final vowel loss process occurs *after* LVD in all cases. Finally, Blevins (2009) expanded the scope of the study of LVD outside of Oceanic (and Austronesian) by bringing into attention the case of synchronic LVD in Alamblak using data from Bruce and others (1984). Alamblak is a Sepik-Hill language with no confirmed genetic relationship to Oceanic languages. Like the previously studied cases, it is the first vowel that is raised in Alamblak. Blevins also mentions processes in a few other languages (Kera, Russian, and certain East Slavic languages) that may count as LVD.

The present article presents a productive form of LVD in Mazandarani. Unlike the previously seen cases, there are two low vowels that take part in providing the environment for LVD in Mazandarani (/æ/ and /ɒ/), but only one of them (/æ/) can undergo raising. This gives rise to more complex patterns in the occurrence of LVD. Moreover, in Mazandarani, it is usually the second — rather than the first — vowel in a pair of consecutive syllables with low vowels that is raised. As we shall see in Section 3.2, this varies depending on dialect.

2. LVD in Mazandarani verbs

We base our discussion on the dialect of Amol — which stands in the middle of those of Babol and Reineh with regard to the features that are of interest to us — and make reference to the other two dialects only when necessary. Unless otherwise stated, all of the vowel change patterns reported in the paper apply in the dialects of Reineh and Babol too (with minor differences that are irrelevant to LVD in certain words). In this section, the only point of difference worthy of mentioning is that in all cases of vowel raising, the resulting vowel is [e] rather than [ə] in Baboli.

There are six vowels in the dialect of Amol: two low vowels (/æ/ and /ɒ/) plus four non-low vowels (/i/, /u/, /e/, and /ə/).[1] For some speakers (presumably those more influenced by Persian), the vowel /o/ shows up too in some loanwords. More conservative speakers replace it with other vowels (/ə/ or /u/). The vowels are shown in (2).

(2)

	Front	**Center**	**Back**
High	i		u
Mid	e	ə	(o)
Low	æ		ɒ

Both of the low vowels are involved in LVD. We begin by examining how adjacency of syllables containing /æ/ and /ɒ/ in the underlying form is handled in verbal morphology. We use the Mazandarani negation verbal prefix to demonstrate the effect of LVD. The unmarked form of the negation prefix is /næ/, used for both past and present verbs as (3) demonstrates.

(3)

	Verb		Negated form	
a.	ˈxərdə	‘was eating’	ˈnæ-xərdə	‘was not eating’
b.	ˈʃurdə	‘was washing’	ˈnæ-ʃurdə	‘was not washing’
c.	ˈgirnə	‘gets’	ˈnæ-jinə	‘does not get’
d.	ˈʃunə	‘goes’	ˈnæ-ʃunə	‘does not go’
e.	ˈdenə	‘gives’	ˈnæ-denə	‘does not give’
f.	ˈdiə	‘was seeing’	ˈnæ-diə	‘was not seeing’
g.	ˈzuə	‘was hitting’	ˈnæ-zuə	‘was not hitting’

[1] For a general survey of vowels in different dialects of Mazandarani, see Borjian (2019).

The examples above are chosen such that verb stems with different non-low vowels (/i u e ə/) as their first vowel are represented. Moreover, the stems in these examples cover all possibilities in terms of the number of consonants following the first vowel: two (examples a to c), one (examples d and e), and zero (examples f and g). The negation prefix and the verb stem both remain intact in all cases as long as the first vowel of the stem is a non-low vowel.
Let us now look at cases where the first vowel of the stem is underlyingly the low vowel /æ/. Vowels that undergo change are marked with underlines in (4).

(4)

	Verb		Negated form	
a.	ˈʃæ̲nəssə	'was spilling'	ˈnæ-ʃə̲nəssə	'was not spilling'
b.	ˈvæ̲rdə	'was carrying'	ˈnæ-və̲rdə	'was not carrying'
c.	ˈkæ̲ʃiə	'was pulling'	ˈnæ-kə̲ʃiə	'was not pulling'
d.	ˈpæ̲ʤənə	'cooks'	ˈnæ-pə̲ʤənə	'does not cook'
e.	ˈzæ̲nnə	'hits'	ˈnæ-zə̲nnə	'does not hit'
f.	ˈvæ̲nnə	'closes'	ˈnæ-və̲nnə	'does not close'

When the first vowel of the stem is /æ/, adding another syllable with the vowel /æ/ to the left creates a sequence of two syllables with low vowels. Our analysis is that in order to avoid this sequence, the second vowel changes to a non-low vowel ([ə]). Unlike the Oceanic cases, it is the second (rather than the first) vowel that is raised.
The examples in (4) only involve /æ/. By bringing the other low vowel of the language (/ɒ/) into the game, things get more complicated. Consider the verbs in (5), in which the first vowel of the verb stem is /ɒ/.

(5)

	Verb		Negated form	
a.	ˈsɒtə	'was building'	ˈnə̲-sɒtə	'was not building'
b.	ˈdɒə	'was giving'	ˈnə̲-dɒə	'was not giving'
c.	ˈkɒʃtə	'was planting'	ˈnə̲-kɒʃtə	'was not planting'
d.	ˈsɒzənə	'builds'	ˈnə̲-sɒzənə	'does not build'
e.	ˈkɒjnə	'plants'	ˈnə̲-kɒjnə	'does not plant'

In these examples, it is the vowel in the verbal prefix itself (i.e. the first vowel in the word) rather than the verb stem that undergoes raising. What these examples suggest — and other cases discussed in the next section confirm — is that even though the vowel /ɒ/ counts as a low vowel in creating the environment for LVD, it never undergoes raising.
The data presented above involved only the negation prefix /næ/. The effect is visible in the same manner in the behavior of other verbal prefixes too,

most notably the prefix /bæ/ that appears behind perfective, subjunctive, and imperative verbs. For instance, from the stem /værd/for 'carry' (row b in 4), we get [bæ-vərd-ə] `she/he/it carried' with the vowel changing in exactly the same manner as we saw in (4). Similarly, the prefix /bæ/ itself undergoes vowel raising when followed by a syllable featuring /ɒ/. For instance, from the stem /sɒt/ (row a in 5), we get [bə-sɒt-ə] 'she/he/it built' in the same manner as we see in (4). For verbs that require the preverb /dæ/ instead of /bæ/, the same phonological change occurs in either the stem or the prefix in the same manner as we see in (4) and (5).

One might argue that the vowel change under discussion may be viewed as vowel reduction or involve a related metrically induced phenomenon. However, we have sufficient reason to rule out this possibility. Mazandarani (like most — perhaps all — Iranian languages) does not have secondary stress. Thus, there is hardly any motivation to assume binary feet of any kind for this language (but see Rahmani 2019 for an attempt to attribute binary feet independent of stress to Persian words). Focusing on main stress alone, we observe that the stress pattern is not related to the vowel alternation in any meaningful way. In all of the verbs we examined, the stress is on the preverb; yet raising targets the preverb in some cases and the stem in others. Moreover, note that the vowel is raised to [e] rather than [ə] in the dialect of Babol. In this case, it is not easy to argue that the target vowel is "reduced". In the next section, it is shown that LVD targets both stressed and unstressed vowels.

3. LVD in Mazandarani loanwords

The vast majority of loanwords in Mazandarani, including the ones that originally come from Arabic or European languages, have entered the language through Persian. Thus, in what follows, we take the Persian forms of the words as their underlying forms. This does not complicate matters since the phonological systems of the two languages are very close.

The vowel /o/ is relatively rare in the dialects of Amol and Reineh (but not Babol), and Persian /o/ is usually replaced with [ə] or [u]. Beside this, the main process of vowel change in loanword adaptation is that the low vowel /æ/ is sometimes replaced by [ə] ([e] in Baboli). We argue that this change must be analyzed differently from what we see in the case of /o/. Unlike /o/, /æ/ is present in the language's vowel inventory. Thus, the driver for changing /æ/ in loanwords cannot be a categorical tendency to avoid this vowel, but to satisfy other context-dependent constraints.

We argue that the vast majority of the cases where a Persian /æ/ changes in loanwords must be analyzed as cases of LVD, functioning in the same manner as what we observed in verbs. Looking at the vowel change as a

manifestation of a phonological constraint against adjacent syllables with low vowels, one can expect there to be a bias against such sequences in the lexicon of the language too. This is indeed confirmed at least tentatively; the authors could not find any native words with adjacent syllables that have low vowels. When it comes to LVD in loanwords, the entirely systematic and exceptionless process that is visible in the native words and the verbal system cannot be observed. However, the power of LVD to account for the cases of vowel raising in loanwords in general is still quite significant.

3.1. Adjacent syllables with non-identical vowels

A list of loanwords with sequences of two adjacent syllables involving both the vowels /æ/ and /ɒ/ in the underlying form is shown in (6). Note that in most of the example sets presented in this section, some of the loanwords are recent, bearing witness to the fact that the process under discussion is still productive in the language. Recall that /o/ changes to [ə] for independent reasons.

(6)

	Persian	Mazandarani	Gloss
a.	xæ̲ˈtɒ	xə̲ˈtɒ	'error'
b.	fæ̲ˈrɒr	fə̲ˈrɒr	'escape'
c.	ɢæ̲ˈtɒr	ɢə̲ˈtɒr	'train'
d.	tæ̲sɒˈdof	tə̲sɒˈdəf	'accident'
e.	mostæ̲ˈfɒ	məstə̲ˈfɒ	(male first name)
f.	mobtæ̲ˈlɒ	məbtə̲ˈlɒ	'afflicted'
g.	ɒˈdæ̲m	ɒˈdə̲m	'person'

As expected, based on what we saw in verbs, it is always the vowel /æ/ that is raised, regardless of the order of the syllables. In all of these examples, the syllable that undergoes raising has at most one coda consonant. Examples with two coda consonants (which is the maximum allowed in Mazandarani) are rare, but in the few examples that the authors could find, LVD does not occur, suggesting that only syllables with fewer coda consonants are susceptible to change: [nɒmærd] "unmanly" and [ɒhæng] "music".

In the examples we have seen so far, the two vowels are separated by only one consonant. The process can occur when consonant clusters separate the two vowels too, as shown in (7). The second example in this list may be viewed as a cognate rather than a loanword, but it helps in showing the effect under discussion nevertheless. As we shall see, the same word appears without raising in the dialect of Reineh. In all of the examples in (7), the first vowel is /æ/ and the second one is /ɒ/. We could not find cases of raising where the

original Persian word features ɒCCæ. However, this may be due to the fact that /ɒ/ is long (VV) in Persian and words with medial VVC syllables are rare in the first place, reflecting a bias in the Persian lexicon disfavoring two coda consonants following long vowels (Samareh 2009 [1999], pp. 146–147).

(7)

	Persian	Mazandarani	Gloss
a.	ʃæ̲lˈvɒr	ʃə̲lˈvɒr	'pants'
b.	hæ̲ʃˈtɒd	hə̲ʃˈtɒd	'eighty'
c.	æ̲rˈvɒh	ə̲rˈvɒ(h)	'souls'

Nevertheless, consonant clusters apparently do make it less likely for LVD to occur, as there is a large number of loanwords of this type where LVD does not occur, e.g. [xæjjɒt] ("tailor"), [ærbɒb] ("master"), [æxlɒɢ] ("behavior"), [pɒjtæxt] ("capital city"), [pɒkdæst] ("incorruptible"). The effect of consonant clusters is more visible when different dialects are compared. This is one of the cases where the dialects we examined seem to behave differently. The dialect of Amol, which is represented in (7), stands somewhere in the middle in terms of how much it favors raising. In the dialect of Reineh, all of the words in (7) occur without vowel raising. In other words, consonant clusters seem to block raising in this variety (more examples of this are presented later when adjacent syllables with the vowel /æ/ are discussed in Section 3.2.). On the other hand, Baboli shows a stronger tendency towards raising in words involving consonant clusters, applying raising in some words that the dialects of Amol and Reineh do not change, e.g. [ɢəssɒb] (cf. Persian [ɢæssɒb] "butcher"), [pəndʒ̄ɒh] (cf. Persian [pændʒ̄ɒh] "fifty"). This is part of a more general trend that we shall see through this work; the dialect of Reineh shows the lowest degree of tendency towards raising while the dialect of Babol is most likely to raise vowels.

To confirm that it is indeed LVD that is responsible for the changes discussed so far, it is necessary to also look at cases where the syllables with low vowels are *not* adjacent to other syllables with low vowels. A list of such words where raising simply does not occur is shown in (8).

(8)

	Persian	Mazandarani	Gloss
a.	ziˈvær	ziˈvær	(female first name)
b.	sæbˈzi	sæbˈzi	'vegetable'
c.	mæˈriz	mæˈriz	'ill'
d.	kæˈbed	kæˈbed	'liver'
e.	æruˈsæk	æruˈsæk	'doll'
f.	mohˈkæm	məhˈkæm	'firm'
g.	muˈʃæk	muˈʃæk	'missile'

There are exceptional cases where raising occurs in such environments too. The most important set of examples is words ending in the (originally Arabic) nominalizing suffix /æt/. The vowel in this suffix is often raised (especially in Baboli), for reasons that are not related to LVD. Examples with this suffix are presented below. Our Amoli speakers pronounced only some of these with raising and did not always agree. The dialect of Reineh does not feature raising in any of these words.

(9)

	Persian	Maz. (Babol)	Gloss
a.	nowˈbæt	noˈbe̲t/nuˈbe̲t	‘turn’
b.	mosiˈbæt	mosiˈbe̲t	‘disaster’
c.	ziˈnæt	ziˈne̲t	(female first name)
d.	suˈræt	suˈre̲t	‘face’
e.	sohˈbæt	suˈbe̲t/sohˈbe̲t	‘conversation’
f.	mosɒfeˈræt	mosɒfeˈre̲t	‘travel’

In addition to these, there are some words in which raising occurs in the absence of the environment for LVD, especially in Baboli and always in the last syllable. A few examples are presented below. We do not have an explanation for these cases, but their restriction to the last syllable does suggest that they involve an effect independent of the phenomenon we are interested in. It must be noted that the last three examples in the list below are words of Iranian origin (the first one is probably of Turkic origin; Hassandoost 2016 [2013], p. 2136). Therefore, at least in theory, rather than viewing them as loanwords, it is possible to view them as cognates or (more plausibly) affected by now-obsolete cognates in their pronunciation.

(10)

	Persian	Maz. (Babol)	Gloss
a.	koˈtæk	keˈtek	‘beating up’
b.	suˈzæn	suˈzen	‘needle’
c.	rowˈʃæn	ruˈʃen/roˈʃen	‘lighted’
d.	dʒ̅iˈgær	dʒ̅iˈger	‘liver’

We argued earlier that vowel raising in Mazandarani is largely independent of stress, citing as evidence the fact that it targets all positions in a word. One might argue that these cases pose a counterexample to our generalization by showing that word-final syllables are indeed special. However, the fact that these words do not involve adjacent syllables with low vowels shows that they are of a different nature from the LVD process we see in verbs and the vast majority of the raising cases in loanwords. In other

words, there is independent motivation to treat these cases as being of a different nature than the main raising phenomenon we are interested in.

We also have good reason to believe that even though stress may have some minimal role in LVD (see Section 4), the data in (10) are not related to stress. When there is interaction between vowel alteration and stress, the cross-linguistic pattern is that reduction (as well as other forms of vowel change) is *prevented* in stressed positions. This is true in known LVD cases that interact with stress too (see Blevins 2009). What we see here is the opposite effect; the exceptional Baboli cases show raising in the final (stressed) positions only. Thus, it is reasonable to assume that the effect we see in these Baboli words is related to word-final position but independent of stress.

3.2. Adjacent syllables with identical vowels

It is now time to look at cases where two adjacent syllables have identical low vowels in the underlying form. When the two vowels are /ɒ/, raising categorically fails to apply.[2]

(11)

	Persian	Mazandarani	Gloss
a.	bɒˈlɒ	bɒˈlɒ	'up'
b.	ɒˈzɒd	ɒˈzɒd	'free'
c.	bɒʃˈgɒh	bɒʃˈgɒh	'club'
d.	ɒmɒˈde	ɒmɒˈdə	'ready'
e.	modɒˈrɒ	mədɒˈrɒ	'tolerance'
f.	ɒgɒˈ(h)i	ɒgɒˈ(h)i	'police station'

The more interesting cases are those in which both vowels in a sequence of syllables in the underlying form are /æ/. In such words, the choice of which syllable to change depends on the dialect. In the speech of our Baboli speakers, it is usually the second vowel that is raised in words of this type (similar to what we saw in verbs). In the dialect of Reineh, however, it is always the first vowel that changes. Our three Amoli speakers were divided in where they apply the raising in such words. Note that since words

[2] Some of the examples in (11) are of Iranian origin. An anonymous reviewer expresses concern over the fact that Mazandarani words of Iranian origin may be cognates rather than loanwords. We believe this is very unlikely in these particular cases based on what we know about the phonology of the two languages and the history of these words. However, even if this is the case, what matters most is that the Mazandarani words in (11) allow sequences of syllables with /ɒ/ in their surface forms.

generally do not end in [æ] in either Persian or Mazandarani, none of the examples have a word-final open syllable.

(12)

	Persian	Maz. (Reineh)	Maz. (Babol)	Gloss
a.	bæˈlæd	bə̲ˈlæd	bæˈle̲d	'knowing'
b.	ɢæˈlæt	ɢə̲ˈlæt	ɢæˈle̲t	'wrong'
c.	hæˈsæn	hə̲ˈsæn	hæˈse̲n	(male first name)
d.	næˈzær	nə̲ˈzær	næˈze̲r	'opinion'
e.	ɢæˈdæm	ɢə̲ˈdæm	ɢæˈde̲m	'stroll'

What triggers the raising is the tendency to prevent two adjacent syllables containing [æ]. The above data show that dialects may vary regarding how they avoid this surface configuration, but they share the active constraints that drive LVD in the first place.

As before, LVD seems to occur with very few exceptions wherever only a single consonant separates the two low vowels. When a consonant cluster comes in between the vowels, LVD does not occur in the dialect of Reineh, but it sometimes does in Amol and Babol. The examples below show the data for Amol. Those of Babol are identical, with [e] instead of [ə] as the raised vowel.

(13)

	Persian	Maz. (Amol)	Gloss
a.	mæɢˈsæ̲d	mæɢˈsə̲d	'destination'
b.	mæxˈzæ̲n	mæxˈzə̲n	'container'
c.	mærˈɢæ̲d	mærˈɢə̲d	'shrine'
d.	pænˈtʃæ̲r	pænˈtʃə̲r	'flat tire'

We may now take a step further and consider cases where more than two syllables are involved in LVD. Let us start with words containing three consecutive syllables with the vowel /æ/ in the underlying form. These cases shed light on the nature of the phenomenon. In such words, in the few examples we could find, it is always the middle syllable that undergoes raising, as shown in (14). Under a constraint-based view, this may be accounted for simply as the option that is most faithful to the underlying form (in terms of the number of changes involved) while avoiding adjacent syllable pairs with low vowels. Note that in (14a), the vowel that is expected to raise is in fact omitted in the dialects of Amol and Reineh. We do not have a method for testing whether raising precedes the deletion (either diachronically or synchronously under a serial account) or not.

(14)

	Persian	Maz. (Amol)	Maz. (Reineh)	Maz. (Babol)	Gloss
a.	mætæˈlæk	mætˈlæk	mætˈlæk	mæteˈlæk	'teasing'
b.	kæmærˈbænd	kæmærˈbænd	kæmærˈbænd	kæmerˈbænd	'belt'
c.	kærgæˈdæn	kærgæˈdæn	kærgəˈdæn	kærgeˈdæn	'rhinoceros'

With the same logic, it comes as no surprise that in æ-æ-ɒ sequences, it is again the vowel in the middle that gets raised (if LVD occurs at all). Examples are presented below. Note that raising occurs only in Baboli for some of these examples. For ɒ-æ-æ, we could not find an example that undergoes a consistent vowel change.

(15)

	Persian	Maz. (Amol and Reineh)	Maz. (Babol)	Gloss
a.	xælæˈbɒn	xæləˈbɒn (only Amol)	xæleˈbɒn	'pilot'
b.	sælæˈvɒt	sæləˈvɒt	sæleˈvɒt	'religious praise'
c.	tæræfˈdɒr	tæræfˈdɒr	tærefˈdɒr	'supporter'
d.	tælæˈfɒt	tælæˈfɒt	teleˈfɒt	'casualties'
e.	dæs(t)ænˈdɒz	dæsənˈdɒz	dæsenˈdɒz	'bump'

To summarize our findings, we present the differences in vowel raising across the three dialects examined in this study in (16).

(16)

	Amol	Reineh	Babol
Raising pattern	æ → ə	æ → ə	æ → e
Preference in æ-æ sequences	(divided)	Raise the first vowel.	Raise the second vowel.
Features word-final raising?	rarely	rarely	occasionally
Features raising in VCCV environments?	rarely	no	occasionally

4. Discussion

There are a number of factors that make LVD in Mazandarani theoretically and typologically interesting. First of all, LVD is a typologically rare phenomenon and little progress has been made in understanding the

articulatory or structural factors that induce it. In fact, the very existence of true vowel dissimilation in human languages has been called into question (see Bennett 2015, Section 1.1). Outside of Oceanic, the cases of LVD identified by Blevins (2009) are limited to Alamblak (Sepik-Hill), several East Slavic language varieties, Kera (Chadic; Ebert 1979), and Wintu (Witnun; Pitkin 1984). Even among these few cases, not all are straightforward cases of LVD. In the East Slavic cases, rather than an underlying low vowel raising to a non-low vowel, the dissimilatory effect manifests itself through a vowel failing to change to [a] in certain environments. In Wintu, the process targets /eCa/ and /oCa/ sequences (meaning that rather than low vowels, it targets non-high vowels), but fails to apply in the case of /aCa/ sequences.

One of the most important aspects of the Mazandarani LVD mechanism is that it often leads to the raising of the second syllable in the sequence. In /ɒCæ/ sequences (e.g. 6g and 6h), this can be explained by the systematic avoidance of altering /ɒ/ (we discuss the reasons for the different behavior of /ɒ/ in the next section). However, in Baboli, even in /æC(C)æ/ sequences, it is the second vowel that undergoes raising (see the examples in 12 and 13). Moreover, in prefixed verbs (but not in nouns and adjectives), the Mazandarani dialects of Amol and Reineh also favor raising the second syllable (see the examples in 3). This is interesting because in almost all other known cases of LVD, it is the first vowel that undergoes raising. The only potential exception according to Blevins (2009) is the Neve'ei (Oceanic), where the suffix /-Vn/, in which the vowel changes shape in harmony with the preceding vowel, fails to appear as [a] after a preceding [a], presumably for dissimilatory reasons (LVD does occur elsewhere in the language too, but targets the first vowel in those cases). While invoking LVD to explain the failure of vowel harmony in such environments in Neve'ei seems reasonable, the effect is less clear than the Mazandarani case. Thus, Mazandarani (especially in its Baboli variety) gives us the only clear example of LVD preferring to raise the second vowel.

There is another aspect of the choice of vowels to raise that is worthy of examination. In the dialect of Babol, raising the second vowel is always preferred. However, in the dialects of Reineh, we observed that while raising targets the first vowel in nouns and adjectives (e.g. /næzær/ 'opinion' appearing as [nəzær]) it targets the second vowel in prefixed verbs (e.g. /næ+værdə/ 'did not carry' appearing as [næ+vərdə]). This may be due to the fact that the first syllable is stressed in prefixed verbs. This is in line with the general cross-linguistic observation that stressed vowels are more stable and the fact that being unstressed is a precondition for undergoing raising in LVD in some other languages too (Lynch 2003, Blevins 2009). However, confirming this hypothesis requires examining a wider range of examples,

e.g. cases where none of the vowels in a /æCæ/ sequence is stressed and there are no low vowels in adjacent syllables. Given the scarcity of such words and the limitations of our elicited data, we leave a thorough examination of the issue for future research.

5. LVD and vowel length

We end this paper with a relatively short discussion on the difference between the two low vowels /ɒ/ and /æ/ in Mazandarani. We observed that even though both of these vowels participate in creating the environment for LVD, it is only /æ/ that can be raised. Further research is needed to arrive at a definitive explanation of this fact, but one particular tentative answer seems to be worth mentioning. It is already well-known in the literature on Persian phonology that the long vowels (/ɒ u i/) are more stable and less susceptible to change in comparison to the short vowels /æ e o/ (Lazard 1957, Toosarvandani 2004). It seems reasonable to argue that their etymological counterparts in Mazandarani, i.e. /ɒ u i/ are long too.[3] We are already aware of the long status (both phonetically and phonologically) of these vowels and their "stability" in the closely related language Gilaki (Rastorgueva et al. 2012 [1971], p. 9).

We do not have access to phonetic evidence to support this and our impressionistic assessment is that duration differences between the two sets of vowels in Mazandarani are either small or non-existent. However, at least at an abstract phonological level, we argue that the vowels /ɒ u i e/ behave as long while the other vowels are short. In this regard, the situation is similar to modern spoken Persian, where most phonetic measurements suggests that duration differences between the so-called "short" and "long" vowels have largely (if not completely) disappeared (e.g. see Moosavi 2011, Sheykh Sang Tajan & Bijankhan 2013, Jones 2019, but also Sadeghi 2013) while phenomena sensitive to phonological vowel length such as versification in this language variety (e.g. in folk poetry) still treat the two vowel classes differently in terms of moraic length (Vahidian Kamyar 1978, Fatemi 2014, Mahdavi Mazdeh 2020). If this is the case in Mazandarani, the permissibility of applying changes to /æ/ (but not /ɒ/) is parallel to the phenomenon observed in Persian by Lazard (1957) wherein only short vowels readily undergo changes. The higher susceptibility of short vowels to change is

[3] The other long vowel in Mazandarani is /e/. From a diachronic perspective, this vowel does not correspond to modern Iranian Persian /e/, but to Early New Persian long /e/ (the vowel traditionally referred to as *yâ-ye majhul*). This vowel has merged with /i/ in modern Iranian Persian. For instance, Mazandarani /ser/ "full" and /ged͡ʒ/ "absent-minded" correspond to the same forms in Early New Persian, but to /sir/ and /gid͡ʒ/ in modern Iranian Persian.

cross-linguistically common and, as pointed out by Blevins (2009), is reflected in known LVD cases too.

One important piece of evidence for the claim that phonological vowel length distinctions may be active in Mazandarani phonological processes comes from the choice of vowels in loanword adaptation. Let us start with the case of /o/ in loanwords. The phenomenon that is of interest to us manifests itself most clearly in the dialects of Amol and Reineh. In these dialects, Persian /o/ is generally replaced with /ə/:

(17)

	Persian	Maz. (Amol and Reineh)	Gloss
a.	mo̲ʃˈkel	mə̲ʃˈkel	'problem'
b.	ɢo̲rˈbun	ɢə̲rˈbun	(male first name)
c.	taʃakˈko̲r	təʃækˈkə̲r	'thanks'
d.	ko̲d	kə̲d	'code'

Crucially, in environments where the vowel is followed by a deleted coda consonant, the vowel replacing /o/ is generally an [u]. Examples are shown below.

(18)

	Persian	Maz. (Amol)	Maz. (Reineh)	Gloss
a.	so̲hˈbæt	su̲ˈbət	səhˈbæt	'conversation'
b.	no̲wˈbæt	nu̲ˈbæt/nu̲ˈbət	nu̲ˈbæt	'turn'
c.	ho̲wl	hu̲l	hu̲l	'fear'
d.	ho̲wseˈle	hu̲sˈlə	hu̲sˈlə	'patience'

This can be accounted for as follows: deleting the consonant (or, under an alternative analysis of cases b to d, the second part of the diphthong) removes a mora. In many languages, when a coda consonant is removed, the missing mora is compensated for by replacing the short (monomoraic) vowel with a long (bimoraic) vowel. This cross-linguistically common process of compensatory lengthening occurs in Persian too (Darzi 1993, Shademan 2005, Sadeghi 2011). We may argue that in Mazandarani, the choice of /u/ instead of /o/ is related to the loss of the consonantal mora. In the words in (18), a long vowel is preferred because it compensates for the missing mora. If this account is correct, it serves as evidence showing that /u/ behaves as a phonologically long vowel in this variety (and probably other varieties) of Mazandarani, while /ə/ behaves as short.

The above discussion suggests that a vowel length distinction is indeed active in Mazandarani phonology. To show that /ɒ/ is long too, we need to find similar cases where /ɒ/ appears in the output when long vowels are

expected. We could find two examples where the sequence /oh/ is rendered as [ɒ] in Mazandarani.

(19)

	Persian	Maz. (Amol)	Maz. (Reineh)	Gloss
a.	moh̲ˈsen	muˈsen	mɒ̲ˈsen	(male first name)
b.	foh̲ʃ	fɒ̲ʃ	fɒ̲ʃ	'profanity'

Even though the two examples above for /ɒ/ are far from adequate, the similarity to the case of /u/ and our prior knowledge of the long status of /ɒ/ in related languages give plausibility to the idea that /ɒ u i e/ are phonologically long in Mazandarani. Thus, we may argue that LVD applies in Mazandrani to prevent adjacent syllables with low vowels, but it can only raise short vowels. From a constraint-based standpoint, this may be justified by assuming that changing two moras is costlier than changing one mora, and (as the data provided in this paper suggest) costlier than having two adjacent syllables with low vowels on the surface.

BIBLIOGRAPHY

Bennett, Wm. G. (2015), *The phonology of consonants*, Cambridge: Cambridge University Press.

Blevins, J. (2009), "Low vowel dissimilation outside of Oceanic: The case of Alamblak", *Oceanic Linguistics*, vol. 48(2), pp. 477–483.

Blust, R. (1996a), "Low vowel dissimilation in Ere", *Oceanic Linguistics,* vol. 35(1), pp. 96–112.

Blust, R. (1996b), "Low vowel dissimilation in Oceanic languages: An addendum", *Oceanic Linguistics,* vol. 35(2), pp. 305–309.

Borjian, H. (2019), "Mazandarani: A typological survey", in: *Essays on Typology of Iranian Languages*, eds. Alireza Korangy & Behrooz Mahmoodi-Bakhtiari, Berlin: De Gruyter, pp. 79–101.

Bruce, L. (1984), *The Alamblak language of Papua New Guinea (East Sepik)*, Canberra: The Australian National University.

Darzi, A. (1993), "Keshesh-e jobrâni-e mosavvethâ dar fârsi-e mohâvere'i-e emruz [Compensatory lengthening of vowels in contemporary Colloquial Persian]", *Zabânshenâsi,* vol. 10(2), pp. 58–75.

Ebert, K. H. (1979), *Sprache und Tradition der Kera (Tschad), Teil III: Grammatik*, Berlin: Dietrich Reimer.

Fatemi, S. (2014), "Rythmique enfantine en Iran", *Revue Des Traditions Musicales Des Mondes Arabe et Méditerranéen*, vol. 8(1), pp. 85–108.

Hassandoost, M. (2016) [2013], *Farhang-e risheshenâkhti-e zabân-e fârsi [Persian etymological dictionary],* vol. 4), Farhangestân-e Zabân-o Adab-e Fârsi.

Jones, T. (2019), "A corpus phonetic study of contemporary Persian vowels in casual speech", *University of Pennsylvania Working Papers in Linguistics*, vol. 25(1), pp. 129–137.

Kord Zaferanloo Kambuzia, A., and Ezatabadipour, T. (2018). "Barresi-e farâyand-e erteqâ-ye vâke va vâzhguni-e in qâ'ede dar gunehâ-ye zabân-e fârsi: ruykard-e zâyeshi [Vowel raising and its rule conversion in varieties of Persian language: A generative phonology approach]", *Motâle'ât-e Zabânhâ va Guyeshhâ-ye Gharb-e Irân,* vol. 5(1), pp. 19–36.

Lazard, G. (1957), *Grammaire du persan contemporain*, Paris: Lib C Klincksieck.

Lynch, J. (2003), "Low vowel dissimilation in Vanuatu languages", *Oceanic Linguistics*, vol. 42(2), pp. 359–406.

Mahdavi Mazdeh, M. (2020), "Quantitative meter in Persian folk songs and pop lyrics", in: *Advances in Iranian linguistics*, eds. Richard Larson & Sedigheh Moradi & Vida Samiian, Amsterdam: John Benjamins, pp. 237–255.

Modarresi Ghavami, G. (2020), "Naqsh-e sâkht-e hejâ dar farâyand-e erteqâ-ye vâke dar zabânhâ va guyeshhâ-ye irâni [The role of syllable structure in vowel raising in Iranian languages and dialects]", *Zabân va Zabânshenâsi*, vol. 16(2), pp. 31–48.

Moosavi, N. (2011), "Barresi-e âkustiki-e vazn-e hejâ dar zabân-e fârsi [An acoustic study of syllable weight in Persian]", *Pazhooheshhâ-ye Zabânshenâsi*, vol. 3(1), pp. 67–76.

Pitkin, H. (1984), *Wintu grammar*, Berkeley: University of California Press.

Rahmani, H. (2019), *An evidence-based new analysis of Persian word prosody*, Amsterdam: LOT.

Rastorgueva, V., Kerimova, A., Mamedzade, A., Pireiko, L., Edel'man, D., and Lockwood, R. M. (2012), *The Gilaki language*, Uppsala: Uppsala University.

Samareh, Y. (2009) [1999], *Âvâshenâsi-e zabân-e fârsi [Persian phonetics]*, Tehran: Markaz-e Nashr-e Dâneshgâhi.

Sadeghi, V. (2011), "Keshesh-e jobrâni dar fârsi: yek âzmâyesh-e edrâki [Compensatory lengthening in Persian: a perception experiment]", *Pazhooheshhâ-Ye Zabâni*, vol. 2(2), pp. 77–94.

Sadeghi, V. (2013), "Barresi-e âvayi-e kâhesh-e vâke'i dar zabân-e fârsi [Phonetic measurement of vowel reduction in Persian]", *Jostârhâ-ye Zabâni,* vol. 6(3), pp. 165–187.

Shademan, Sh. (2005), "Glottal-deletion and compensatory lengthening in Farsi", *UCLA Working Papers in Phonetics*, vol. 104, pp. 61–68.

Shahidi, M. (2008), *A sociolinguistic study of language shift in Mazandaran*, Uppsala: Uppsala University.

Sheykh Sang Tajan, Sh., and Bijankhan, M. (2013), "Barresi-e mâhiat-e sowtshenâkhti-e vâkehâ-ye fârsi dar hejâhâ-ye CV-e goftâr-e ezhâri [A study of the phonetic nature of Persian vowels in CV syllables in citation form]", *Zabânpazhoohi*, vol. 4(1), pp. 97–116.

Toosarvandani, M. D. (2004), "Vowel length in modern Farsi", *Journal of the Royal Asiatic Society*, vol. 14(3). pp. 241–251.

Vahidian Kamyar, T. (1978), *Barresi-e vazn-e she'r-e âmiâne [A study of meter in folk poetry]*, Tehran: Âgâh.