ÖAW ÖSTERREICHISCHE AKADEMIE DER WISSENSCHAFTEN

Ninth International Conference on Iranian Linguistics (ICIL 9)

Booklet of abstracts

December 2020

partner institutions:

CeRMI
CENTRE de RECHERCHE sur le MONDE IRANIEN
UMR 8041 CNRS. SORBONNE NOUVELLE. INALCO. EPHE

ÖAW ÖSTERREICHISCHE AKADEMIE DER WISSENSCHAFTEN

Ninth International Conference on Iranian Linguistics
(ICIL 9)

Institut für Iranistik, Österreichische Akademie der Wissenschaften, Vienna / Austria

initially planned for 20-22 August 2020

keynote speakers:

Jila Ghomeshi (University of Manitoba, Winnipeg / Canada)
Adriano Rossi (Università degli Studi di Napoli "L'Orientale", Naples / Italy)

organisers:

Univ.-Dozent Dr. Velizar Sadovski (Österreichische Akademie der Wissenschaften, Vienna)
PD Dr. Agnes Korn (CeRMI ; CNRS, Paris)
Dr. Oleg Belyaev (Lomonosov State University, Moscow)

Local organisation:

Hon.-Prof. Dr. Florian Schwarz (Österreichische Akademie der Wissenschaften, Vienna)
Univ.-Dozent Dr. Velizar Sadovski (Österreichische Akademie der Wissenschaften, Vienna)
Bettina Hofleitner, BA (Secretary)

partner institutions:

CeRMI
CENTRE de RECHERCHE sur le MONDE IRANIEN
UMR 8041 CNRS. SORBONNE NOUVELLE. INALCO. EPHE

Accepted papers:

Ahmadreza Qaemmaqami	*Old Persian* hucâra-
Ali Hussain Birahimani	*Some consequences of lenition on the verb system of Eastern Balochi*
Anton Zykov	*Parsi (Parsi Gujarati) and Gavruni (Zoroastrian Dari): Language Change and Ritual Lexicon*
Chams Bernard	*On Iranian borrowings in Tocharian*
Daniel Barry	*The Turkic and Semitic sources of /q/ in Kurmanji Kurdish*
Dorian Pastor	*Persian of Tehran: towards a new rise of simple verbs?*
Emine Şahingöz	*Digor-Ossetic Definite Particle i*
Erik Anonby, Bettina Leitner, Mortaza Taheri-Ardali & Dina El Zarka	*Arabic of south Iran: The dialect of Bandar Moqam in Hormozgan Province*
Erik Anonby	*Independent outlier: On the linguistic history of Kumzari*
Fahimeh Tasalli Bakhsh	*Grammaticalization of Modals in Persian*
Fati Antadze	*On One Peculiarity of Lexicalization of Arabic Broken Plural in Persian*
Gerardo Barbera	*Bashkardi and the languages of Southeast Iran*
Harald Bichlmeier & Maria Carmela Benvenuto	*The expression of predicative possession in Avestan*
Jeremy Bradley, Veronika Milanova, Sampsa Holopainen & András Czentnár	*Structural parallelism in Turkic, Iranian, and Uralic languages from a contact-linguistic perspective*
Juan Briceño-Villalobos	*On Old Persian* tuvam ka haya *"you, whoever"*
Julian Kreidl	*The reconstruction of Early New Pashto*
Kateryna Iefremenko	*Placement of the locational goal elements in Kurmanji Kurdish in contact with Turkish and German*
Liubov Silanteva	*Multielement Structure Of Wakhi Stress In Pamir Region*
Mahjoub Zirak & Faryar Akhlaghi	*Areal Typological Study of Contact-Induced Features in Arabic Varieties of Iran*
Mohammad Rasekh-Mahand	*The emergence of periphrastic aspectual verbs in Persian*
Mohsen Mahdavi Mazdeh	*Phonological Effects in the Ordering of Persian Binominals*
Murad Suleymanov	*Possessor-Source Syncretism in Abşeron Tat*
Natalia Serdobolskaya	*Differential Object Marking in Iron Ossetic*
Oleg Belyaev & Irina Khomchenkova	*"Ablative coordination" in Ossetic?*
Raha Musavi	*Dashtaki Verb Inflection*
Ronald Kim	*Towards a historical phonology of Ossetic: relative chronology and Alanic evidence*
Saeid Norouzi & Hassan Rafiey	*The Features of the "Iranian Drug Addict's Slang" (IDAS)*
Sara Belelli	*Verbal particles in Laki: Patterns of Polysemy and Development of an Elusive Lexical Class*
Setayesh Dashti	*On the Developments of Bayestan*
Shuan Karim	*The Proto-Zaza-Gorani case system: evidence for convergence and expansion*
Vahid Sadeghi	*On the phonetic differences between the intonation of Persian statements and Questions*
Zahra Etebari	*The rise of contrastive-partitive and definite functions of* =eš *in colloquial Persian: A grammaticalization from third person clitic pronoun*

Posters:

p. 67

Artem Trofimov	*The accentual system of Parachi and Ormuri and its role for placing them among Iranian languages*
A. Ali Ahangar & Setareh Mojahedi Rezaeian	*The Representation of Switch-Participant Reference in Persian-speaking Children's Narrative Discourse*
Agnès Lenepveu-Hotz	*Hesitations in the first French words borrowed by New Persian: The example of the first Safarname-ye Naser ad-din Shah (1873)*
Arseniy Vydrin	*Tajik National Corpus*
CV Anonby	*Said, Seen, or Thought: Evidentiality in Kumzari*
Faryar Akhlaghi	*Adverbial Relations in Modern Persian language: A Typological Study*
Dinara Stepina	*Optative or Subjunctive? Two mood systems of Azerbaijani Talysh*
Jens Helfer-Fleischhauer	*Horizontal alignment and ense/aspect-restricted differential object marking in Iranian languages*
Katarzyna Marszalek-Kowalewska	*'kampyutere shaxsi' or 'rayane shaxsi': Deep Learning Methods for the Extraction of Persian Multi-word Expressions with English Loanwords and their Persian Equivalents*
Maryam Nourzaei	*Definiteness markers in New Western Iranian languages, from a typology perspective*
Mehrnoosh Taherkhani & Farnoosh Taherkhani	*Hybrid alignment patterns in ergative system of Iranian languages: evidence from Takestani*
Minoo Nassajian & Razieh Shojaei	*A Dependency Grammar Approach to Persian Ezafe Construction*
Mohammad Rasekh-Mahand & Elham Izadi	*The corpus-based study post-predicate elements in spoken Persian*
Mortaza Taheri-Ardali	*Definiteness marking in the languages of Chahar Mahal va Bakhtiari Province, Iran*
Muhammed Ourang	*The derivation of light verbs in Aheli*
Nafiseh Tadayyon & Javid Fereidoni	*Code-switching Pattern in Persian: A Sociolinguistic Study in Urmia*
Sampsa Holopainen & András Czentnár	*On some problematic Ossetic etymologies*
Sayuri Igarashi	*The Function of The Passive Voice as An Alternative Expression in Modern Persian*
Shadi Davari & Kian Karbalaei-Mohammadi	*Phrasal and Clausal Superior Comparative Construction in Persian*
Sima Zolfaghari	*Lexical Semantic Fields in Iranian Languages: The case of Bakhtiari Body Parts Terminology*
Songül Gündoğdu, Arsalan Kahnemuyipour & Andrew Peters	*Traces of Genitive Case in Southern Zazaki Ezafe*
Tatiana Kononova	*Building a Tagged Corpus of the Classical Persian Langugage*

Papers

Old Persian *hucāra-*

SEYYED AHMAD REZA QAEMMAQAMI
Ancient Iranian Languages, Faculty of Letters and Humanities, University of Tehran, Iran

The word *hucāra-* occurs four times in Old Persian Inscriptions:
1. DSj 4: mām AM dauštā āha^{t}. taya akunavam avamai visam ^{h}ucāram āha^{t}.
2. DSl 5: θāti D. xšāyaθya vašnā A^{h}uramazdāha taya amanyai kunavānai ava-mai visam ^{h}ucāram āha^{t}.
3. DB 4.76: A^{h}uramazdā θvām dauštā byā^{t} uta-tai taumā vasai byā^{t} uta dargam jīvā^{h} uta taya kunavāhi ava-tai A^{h}uramazdā ^{h}ucāram kunautu.
4. DSf 20: A^{h}uramazdā-mai upastām abarat. tayamai framātam cartanai ava^{t} ^{h}ucāram akunauš.

This word is commonly considered as a *bahuvrīhi* compound, meaning successful and its equivalents. R. Kent (1950: 173) has compared the second member with Av. *cāra-* and NPers. *cāra* and translated it to "well done, successful". Before him, E. Benveniste (1931: 168) has translated this word to "bien réussi, réussit" as well. W. Brandenstein and M. Mayrhofer (1964: 147-8) have also rendered it to "geglückt, gut gelungen, erfolgreich" and suggested that this word must be a *bahuvrīhi* compound orginially meaning "dessen Mittel gut sind". W. Hinz (1973: 138) has turned it to "wohlgelungen, geglückt, erfolgreich" similarly. Eventually, R. Schmitt (1991: 72; 2014: 263) has translated it to "successful" in his translation of Bistun inscription and to "erfolgreich" in his dictionary of Old Persian and believed it to supposedly be a *bahuvrīhi* compound consisting of *hu-* "gut" and *cārǎ-* "Mittel, Verfahren".

The context of the aforementioned OP phrases nonetheless suggests another meaning for this word: "simple, easy", which fits very well in all four phrases and besides is not incompatible with Akkadian and Elamite versions.. Moreover, its continuation in Middle Persian must be *huzārag*. In fact, this word is common in Zarathustrian and Manichean Middle Persian and means "little, few, scant", but no one has pointed out to its relation with OP. *hucāra-*. As it happens, in one or two instances from Middle Persian, *huzārag* has a meaning similar to "easy, simple" in one instance or "cheap, low-grade" in another, and the semantic change of the words meaning "simple, easy" to "little, few, scant" and "cheap, low-grade" is a common change; e.g. Pers. *x^{w}ār* or Ar. *yasīr* which has both meanings and in translations of Qur'an to Early New Persian has been occasionally translated to *xujāra*, a dialectic remnant of MP. *huzārag*.

I suggest that OP. *hucāra-* is not a *bahuvrīhi* compound, rather a verbal or synthetic compound whose second member is not the noun *cāra-* but a verbal component which, similar to a number of Vedic compounds formed with *su-* and *duṣ-*, means nearly "easy to …" and can be compared to e.g. Skt. *su-kára-* "easy to do, easy to accomplish", Skt. *duš-cyavaná-* "hard to overthrow, immovable" and Skt. su-yúj- "well-yoked".

Thus, in my opinion, OP. *hucāra-* is
1. the origin of MP. *huzārag* and NPers. *xujāra*,
2. in OP inscriptions means "rendered easy" > "easy" and
3. is not a *bahuvrīhi* compound and hence one has to reconsider the assessment of former researchers.
Analyzing the occurances of this word in OP and its later variants in MP, this research attempts to shed light on its meaning and morphological structure.

References

Benveniste, É. (1931). Grammaire du vieux perse d'A. Meillet, 2e éd. entièrement corrigée et augmentée, Paris. H. Champion.

Brandenstein, W., & Mayrhofer, M. (1964). Handbuch des Altpersischen. Wiesbaden. Otto Harrassowitz.

Hinz, W. (1973). Neue Wege im Altpersischen. Wiesbaden. Otto Harrassowitz.

Kent, R. G. (1950). Old Persian, Grammar, Texts, Lexicon. New Haven. American Oriental Society.

Schmitt, R. (1991). The Bistun Inscriptions of Darius the Great Old Persian text. London. Corpus inscriptionum iranicarum by School of Oriental and African studies.

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

Some consequences of lenition on the verb system of Eastern Balochi

ALI H. BIRAHIMANI
Shanghai Jiao Tong University

The process of lenition in Eastern Balochi (EBal), which uniformly turned singleton stops and affricates in the post-vocalic position to fricatives, sets it apart from the non-eastern dialects. While its importance as the defining characteristic of EBal has been noted for long, its effects other than the phonological ones have not been explored. The data discussed below is based on the author's native competence and the material collected over time, except where referenced otherwise.

Central to verb system of Balochi, as evidence from Western Balochi (WBal) dialects indicates, are the three verbal prefixes, viz. prohibitive prefix *ma-*, negative prefix *na-*, modal prefix *bi-*, and the verbal element *a=/=a* (see Jahani and Korn 2009: 665-666). These morphemes, it is argued in this paper, also participated in lenition, i.e. spirantizing the initial consonant of the verb to which they attached, and ultimately had far-reaching effects. Using the verb *kʰanaɣ* 'to do' as an example, it may be hypothesized how these prefixes would have interacted with the verb through lenition and led to alternations in morpho-syntax, e.g. present-future **man kʰanan* 'I may/will do' but negation of present-future **man na-x·anan* 'I may/will not do', imperative **bi-x·an* 'do!', prohibitive **ma-x·an* 'do not!', etc; similarly, preterite **man kʰuθ* 'I did' but past subjunctive **man bi-x·uθēn* 'should I have done'. Synchronic variation in spoken EBal, e.g. *na-xanã̄* '(I) do/will not do', *na-vīã̄ '(I) do/will not become', na-ɣirã̄* '(I) do/will not take', *ma-xan* 'don't do!', *ma-vay* 'don't be!', *ma-ɣir* 'don't take!' etc. from *kʰan-*, *bī-*, *gir-* respectively, counts as the primary evidence for the fact that verbal prefixes did have a leniting effect on verbs.

Such alternations, found wherever an author has been receptive to variation, is also reflected in the early sources on Balochi, e.g. "*jar hechi ma vīθī*" 'clothes none must he have on' (Lewis 1885: 7) or "*tavivā durāh na vī*" 'by the doctor (he) cannot be cured' (Lewis 1885: 13), where the initial of *bī-* has been lenited to *vī-*. Elfenbein (1966: 11) noticed "many cases of sentence-sandhi..." and offered as examples "*ā lēṛavā na ðārī*" 'he does not possess a camel' with *dār-* 'hold' and "*taī dilgir ma vī*" '(let) your heart become not alarmed' with *bī-*, and affirmed as late as the collection of Bashir's (2008: 54) data "…*guḍḍeθa na xanã̄*" '(I) cannot cut down' with *kʰan-*.

From the earliest sources it can be ascertained that post-vocalic lenition stopped operating long ago, because these already have words - i.e. loanwords of Perso-Arabic or Indo-Aryan provenance mostly - with post-vocalic singleton stops, e.g. "*khatōla*" 'bedstead', "*phaṭ*" 'wound', "*rich*" 'bear', "*dāchī*" 'female camel', in Leech (1838: 610-611). Such borrowings after spirantisation ceased, created a potential phonemic opposition between stops and fricatives, as Morgenstierne (1948: 256) has already shown, even if only a positional contrast. This build-up of stop-fricative contrast had an important implication for the verb system, that is, the alternations with the verbal prefixes, e.g. **man kʰanan*, **man na-xanan*, **bi-xan*, **ma-xan*, would be grammaticalised eventually; such a result is observed cross-linguistically, for instance, in the initial consonant mutations in Celtic languages where lenition exists as a grammatical phenomenon (Hickey 1997). In EBal context another possibility was that the alternations would be rejected (somehow).

Interestingly, subsequent developments reveal a tendency to rid the system of these alternations, which explain why they could not be grammaticalised into productive rules of morpho-syntax. For example, *ma-* and *na-* survive as prefixes primarily with the verbs beginning with vowels or sonorants, e.g. *ma-ill* 'don't let!' and *na-illay* '(you) do not let', *ma-rav* 'don't go!' and *na-ravay* '(you) do not go' or relics with the lenited forms of the verb. These prefixes have otherwise gone through an array of changes that lead us to describe them

as "degrammaticalised", such as in being put after the verb, e.g. "*ak^{h}is ma* 'don't delay!'" (literally 'don't sleep!') (Leech 1838: 613) for *ma-ak^{h}is* and "*k^{h}as ǰōrēnī na*" 'nobody builds (it)' (Lewis 1885: 3) for *k^{h}as na-ǰōrēnī*. They may appear before the verb, but lengthened into independent "particles" [mā] and [nā], e.g. *kār mā k^{h}an* 'do not work!' and *kār nā k^{h}uθa*(*γ*) 'did not work', or at times cliticized onto the preceding word, e.g. *kār=ma k^{h}an* 'do not work!' and *kār=na k^{h}uθa*(*γ*) 'did not work', the following verb found in either cases in non-lenited form, apparently due to an effect of word-boundary perception. Additionally, *na* may also replace *ma-* and supply the prohibitive mood; for instance, Lewis (1885: 2) "*munjhā ma vī*" 'don't be sad!' but also "*munjhā bī na*" 'don't be sad!' (Lewis 1885: 7).

The prefix *bi-* is largely absent from EBal, and as Dames (1907: 184) noted long ago, it occurs only with the verbs beginning with vowels and *r* and *w*, e.g. *bi-yā* 'come!' *ā-*, *bi-yār* 'bring!' *ār-*, *b-ill* 'let!' *ill-*, *b-ōšt* 'stand!' *ōšt-*, *ba-war*/*b-ōr* 'eat!' *war-*, *ba-rav* 'go!' *rav-*, etc, but otherwise *bay* 'become!', *k^{h}an* 'do!' *k^{h}an-*, *gir* 'take!' *gir-*, ending up as a clitic. In the subjunctive mood too it has a similar distribution, e.g. *bi-yārã̄* '(I) may bring', *b-illã̄* '(I) may let', *ba-ravã̄* '(I) may go', *ba-warã̄*/*b-ōrã̄* '(I) may eat', but also *warã̄* '(I) may eat', *ravã̄* '(I) may go', *bīã̄* '(I) may become', *k^{h}anã̄* '(I) may do', *girã̄* '(I) may take'. The same is also true of the past subjunctive.

The distinctions supplied by *a=*/*=a*, i.e. indicative mood with present-future stem and imperfect tense with past stem, have no trace in EBal. In one set of EBal varieties, the original past stem, e.g. **man k^{h}uθ* 'I did', has taken over the function of a past habitual, i.e. *k^{h}uθ* 'used to do' as already identified by Dames (1922: 25); Korn (2009: 49) notes how a simple past does not occur in Dames (1922) and Gilbertson (1923). The rest of the EBal varieties have a secondary past habitual, i.e. *k^{h}uθaθ* 'used to do' (Dames 1922: 25), whose form though is familiar from WBal dialects analyzed recently by Nourzaei and Jahani (2012). Within the early sources one finds an innovative - as first pointed out by Elfenbein (1982: 88) - periphrastic construction, based on infinitive in the locative case, *-aγ-ā* + COP, e.g. *mã̄ k^{h}anaγā ã̄* 'I am doing'. However, often found attached as a particle with no semantic force, to nouns preceding the verb rather than the verb itself, is *=a*, e.g. *mã̄ kār=a k^{h}anã̄* same in meaning as *mã̄ kār k^{h}anã̄*, *a=*/*=a* too therefore having been "degrammaticalised".

A major outcome of these changes accumulating slowly over time but crucially, it is proposed, initiated to avoid grammatical lenition, was the abolition of the old tense-based system in EBal, which began, it is asserted furthermore, by the "degrammaticalisation" of the verbal element, that had in all likelihood a proclitic attachment originally, *a=*. This verb system thus evolved into one with aspect at the center, like the verb system of New Indo-Ayran languages towards EBal's east.

References:

• Bashir, Elena. (2008). 'Some transitional features of Eastern Balochi: An areal and diachronic perspective'. In Carina Jahani, Agnes Korn, & Paul Titus, (eds.). *The Baloch and Others: Linguistic, Historical and Socio-Political Perspectives on Pluralism in Balochistan* (pp. 45-82). Wiesbaden: Reichert Verlag. • Dames, Mansel Longworth (1907). *Popular Poetry of the Baloches*. Vol. II. London: David Nutt. • Dames, Mansel Longworth (1922). *A textbook of the Balochi language, consisting of Miscellaneous Stories, Legends, Poems, and a Balochi-English Vocabulary.* Lahore: Government Print of Punjab. • Elfenbein, Josef (1966). *The Balochi Language: A Dialectology with Texts*. London: The Royal Asiatic Society of Great Britain and Ireland. • Elfenbein, Josef (1982). 'Notes on the Balochi-Brahui Linguistic Commensality'. Transactions of the Philological Society (pp. 77-98). • Gilbertson, George W. (1923). *The Balochi Language: A grammar and manual*. Hertford. • Hickey, Raymond 1997. Sound change and typological shift: Initial mutation in Celtic. In Jacek Fisiak (ed.), *Linguistic Typology and Reconstruction*, 133-182. Berlin: Mouton de Gruyter. • Jahani, Carina and Agnes Korn 2009. Balochi. In Gernot Windfuhr (ed.), *The Iranian Languages*, 634-692. London & New York: Routeledge. • Korn, Agnes (2009). 'The Ergative System in Balochi from a Typological Perspective.' *Iranian Journal of Applied Language Studies*: 1. pp.43-79. • Leech, R. (1838). 'Grammar of the Balochky Language'. In *Journal of the Bombay Branch of the Royal Asiatic Society of Bengal*. 7:2, pp. 608-620. Calcutta: Asiatic Society of Bengal. • Lewis, Rev. Arthur 1885. *Balochi Stories*. Allahabad: Allahabad Mission Press. • Morgenstierne, Georg 1948. Balochi Miscellanea. *Acta Orientalia*. Leiden. Vol XX. 253-292. • Nourzaei, Mariam and Carina Jahani 2012. The Distribution and Role of the Verb Clitic =a/a= in Different Balochi Dialects. In Geoffrey Haig and Carina Jahani (eds.) *Orientalia Suecana*. Vol LXI. 170-186.

Parsi (Parsi Gujarati) and Gavruni (Zoroastrian Dari): Language Change and Ritual Lexicon

ANTON ZYKOV-GENKE
INaLCO, Paris / University of Frankfurt a.M.

The paper will focus on the ritual lexicon in contemporary Zoroastrian languages of India and Iran, Parsi Gujarati (PG) and Gavruni (aka Zoroastrian Dari). It will explore the link between various religious, historical and sociolinguistic factors that occurred in the 15th – 18th centuries and their influence on the PG ritual lexicon. Data are from Lüddeckens & Karanjia's (2011), Stausberg (2004), as well as my own extensive audio- and video corpus collected in Gujarat and Maharashtra since 2018.

The presentation will employ the research methodology used by Molčanova (2017) to argue that the shift towards Indo-Aryan (IA)-derived lexical items such as in Table 1 describing Zoroastrian purity and pollution rituals among the Parsis was caused by a combination of three factors:

a. the weakened contracts between the two communities,
b. the alternations occurred in the rituals themselves,
c. the influence of the surrounding socio-political and linguistic environment, i.e. sanskritisation in case of PG and persianisation in case of Gavruni.

Table 1 : Ritual terms in PG and Gavruni

	PG	Gavruni
fire-temple	*āgiārī*	*āteškade*
fine open building facing a funerary tower	*bunglī*	*pesgæm*
corpse-bearer outside of *dak^{h}ma* (funerary tower)	*k$^{(h)}$āndiā*	*hamāl*
morsel of consecrated food given to the ritual-performing dog	*kutrānu būk*	*čom-e šwa*
unconsecrated bull urine	*taro*	*pājō*

The paper will also point to a number of linguistic features which differentiate PG from Standard Gujarati (StG) and which are atypical for IA in general, but common for Iranian.

Phonological:

a. dentalisation, hence lack of contrast between dental and retroflex plosives, the latter being a wide-spread characteristic of IA, but normally not found in Iranian (with the exception of Pashto and Balochi), e.g.

StG (Doctor 2004, 9)	PG (own fieldwork)
*pa**ṇ**ī* 'water'	*emā pa**n**ī joīe* '[if] it needs water'

b. de-aspiration: StG aspirated voiceless consonants [p^{h}, t^{h}, ʈ^{h}, tʃ^{h}, k^{h}] are pronounced unaspirated [p, t, ʈ, tʃ, k] and sometimes also thus represented in writing, e.g.

StG (Singh 2017, title)	PG (own fieldwork)
*Hāra pa**c^{h}**ī ja jīta **c^{h}**e*	*Pa**č**ī dokmānu kām čālu t^{h}āī **č**e*
'Only after defeat [there] is victory'	'Afterwards the work [in] *dak^{h}ma* has started'

Morphosyntactic:

a. StG 3PL *evaṇ* (PG *evan*), as HON 3SG (Taylor 1908, 33, also attested in Gujarati Lexicon[1]), cf. Persian 3PL *išān/ešun* (Windfuhr & Perry 2009, 435), Gavruni *udɛm* (Farudi & Toosarvandi 2004, 17).

b. Modi (2011, 54) points out that PG often 'incorrectly' assigns gender, the grammatical category absent in both Persian and Gavruni.

These contemporary speech examples point to a continued influence of Iranian on PG. At the same time, while the existing scholarly view postulates that PG was coined only in the late half of the 19th (JamaspAsa 2002, 392), IA-derived lexical items, with some of them mentioned in Table 1, appear in PG already in the 15th – 18th century sources.

My presentation will explore lexical (mainly focusing on the ritual items), phonological and morphosyntactic change in PG and will make an attempt to provide a plausible explanation of this time discrepancy based on a variety of sources on PG and Gavruni, historical as well contemporary.

This paper will endeavour to provide more evidence on the Parsis' early linguistic history in India while contributing to the understanding on the time when and how the community has shifted from an Iranian to an IA language.

References

Doctor, Raimond. 2004. *A Grammar of Gujarati.* München: Lincom Europa.

Gajendragadkar, S. N. 1974. *Parsi-Gujarati: A Descriptive Analysis*. University of Bombay.

Farudi, Annahita, and Maziyar Toosarvandani. 2004. 'The Dari Language Project: Fieldwork Endeavor. Summary of Findings'. Berkeley: UC Berkeley/linguistics.berkeley.edu.

JamaspAsa, Keikhosrow. 2002. 'Gujarati,' *Encyclopaedia Iranica*, Vol. XI, Fasc. 4, pp. 391-392, available online at http://www.iranicaonline.org/articles/gujarati (accessed online on 17 November 2019).

Lüddeckens, Dorothea, and Ramiyar Karanjia. 2011. *Days of Transition: The Parsi Death Rituals*. Arbeiten Am Lichtenberg-Kolleg. Göttingen: Wallstein Verlag.

Modi, Bharati. 2011. *Parsi Gujarati - Vanishing Dialect: Vanishing Culture.* München: Lincom Europa.

Molčanova, Elena K. 2017. 'Dinamika pogrebal'noj leksiki u sovremennyh zoroastrijcev Irana (Dynamics of funerary lexicon of contemporary Iranian Zoroastrians)'. In *Zur lichten Heimat: Studien zu Manichäismus, Iranistik und Zentralasienkunde im Gedenken an Werner Sundermann*, edited by Alberto Cantera and Maria Macuch, 433–44. Iranica 25. Wiesbaden: Harrassowitz Verlag.

Skjærvø, Prods Oktor. 2016. "Middle West Iranian." In *The Iranian Languages*, edited by Gernot Windfuhr, 196–279. London: Routledge

Stausberg, Michael, ed. 2004. *Zoroastrian Rituals in Context*. Brill.

Taylor, George Pritchard. 1908. *The Student's Gujarati Grammar: With Exercises and Vocabulary*. Bombay: Thacker & Company, Limited.

Windfuhr, Gernot & John R. Perry. 2016. "Middle West Iranian." In *The Iranian Languages*, edited by Gernot Windfuhr, 416–545. London: Routledge

[1] 'evaṇ' In *Gujarati Lexicon*. Accessed 12 December 2019. https://www.gujaratilexicon.com/dictionary/gujarati-to-english-translation/%e0%aa%8f%e0%aa%b5%e0%aa%a3/.

Arabic of south Iran: The dialect of Bandar Moqām in Hormozgan Province

BETTINA LEITNER[1], ERIK ANONBY[2], MORTAZA TAHERI-ARDALI[3], DINA EL ZARKA[4]
[1]University of Vienna; [2]Carleton University; [3]Shahrekord University; [4]University of Graz

Arabic has been spoken as a mother tongue in Iran since pre-Islamic times (Zarrinkūb 1975, Daniel 1986, Nadjmabadi 2009). Arabic in Khuzestan Province (Ingham 1973, Shabibi 2006, Grigore 2011) and Khorasan (Seeger 2002, 2013, Dahlgren 2005) have received some scholarly attention, but other Arabic dialects scattered across the south-west rim of the country – in Ilam, Fars, Bushehr, and Hormozgan – have not been documented.

In this study, we provide a first account of Arabic on Iran's southern coast, with a description of the dialect of Bandar Moqām in western Hormozgan Province. Using the Atlas of the Languages of Iran (ALI) linguistic data questionnaire as well as supplementary elicitation and oral texts, we have documented salient elements of its lexicon, phonology and morphosyntax.

In the lexicon, Bandar Moqām Arabic (BMA) shares most items with Classical Arabic and contemporary Arabic dialects. Geographically, the lexicon falls into the Gulf Arabic (GA) dialect area, and in relation to the B type ("Bedouin") vs. S type ("Sedentary") distinction (Fischer and Jastrow 1980, Holes 2006, 2018), exhibits vocabulary typical of "Bedouin" type dialects (e.g., BMA *ḥərṭəm 'lip', bāčər 'tomorrow', xāf 'maybe', xašəm 'nose'; cf. Rosenhouse 2006*). Distinctive lexical items include *kənīnaw* 'small', otherwise unknown from Arabic but attested in South Arabian languages (Morris et al. in press), and *čība* 'cat', which has not been reported elsewhere. Persian borrowings occur, but in the data we have collected they are generally restricted to nouns (e.g., *panǰera* 'window', *serka* 'vinegar', *xətkār* 'ballpoint pen').

In the phonology, BMA exhibits some typical B type sound changes: Old Arabic (OA) *k* > *č* beside front vowels (*čalb* 'dog', *səmač* 'fish'); OA *q* > *g* (*galḥ* 'heart', *dagg* 'he hit') and sometimes *ǰ* (*ḥalǰ* 'mouth', *ǰatal* 'he killed'); the sound change OA *ǰ* > *y* applies variably (*tənsəǰ* 'she weaves', *yədīd* 'new', *wāyəd~wāǰəd 'a lot'*). Spirantization of OA interdentals *ṯ* (*əsyāb* 'clothes'), *ḏ* (*zān* 'ear') and *ḏ̣* (*ẓəhəṛ* 'noon'), as well as OA *ḍ* (*abyaẓ* < *abyaḍ 'white'*) shows the alignment of BMA with peripheral Arabic dialects such as Central Asian Arabic – and with Persian treatment of Arabic loanwords. Contrastive emphatic consonants are not limited to the Classical/Standard Arabic set *ṭ ṣ ḍ ḏ̣* (with reflexes for the latter two as *ẓ* in BMA), but through historical diffusion in the language have come to include other consonants (*məxaṃṃa* 'broom', *məṛṛ* 'bitter', *ṛəkḅa* 'knee', *galḥ* 'heart'). The phonemes *p* (*parda* 'curtain') and *q* (*barq* 'electricity'; cf. inherited *barg* 'lightning') are found in loanwords from Persian. Vowel-related phenomena of interest are as follows: the distinction between long high vowels, long mid vowels, and diphthongs (*ī*/*ē*/*ay*; *ū*/*ō*/*aw*) is absent word-finally; the historical distinction between short vowels *i* and *u* has been neutralized, and this vowel is frequently inserted epenthetically (*čabəd* 'liver'); word-final *a* is raised to [ɛ] in non-emphatic contexts.

In morphosyntax, BMA again shows mixed patterning in relation to the B type vs. S type distinction as well in relation to geographically-based dialectal divisions between Gulf, Mesopotamian and Central Asian Arabic. The BMA 2FSG pronominal suffix *-əš* as well as the affix *-in*(*n*)-, the latter of which is used between active participles with verbal force and suffixed object pronouns, are typical S type features among Gulf Arabic dialects. This distribution supports our theory of a historical relation or common origin of BMA and the S type dialects of the Gulf. The use of a *b*-prefix for proximate intent is a feature shared with all

contemporary Gulf Arabic dialects. In contrast, in its use of the indefinite article *fad*, BMA patterns with all Iraqi dialects, Khuzestan Arabic, other Arabic dialects in Iran (Dahlgren 2005) and Central Asian Arabic (Seeger 2009), but differently from all other GA dialects described in the literature.

In summary, our analysis confirms that Bandar Moqām Arabic fits into the wider Gulf Arabic dialect area, yet is internally heterogenous. On the one hand, it shares many features with the regionally dominant "Bedouin" type Gulf Arabic koine, but it also aligns with distinctive structures – both retentions and innovations – in the more ancient "Sedentary" Gulf dialects that originated in southern Arabia. Further, it exhibits a series of structures in common with Arabic varieties of Iraq and Khuzestan and its descendants that reached as far as Central Asia during the Arab conquest of greater Persia, and are still spoken there today. Some Persian influence can be ascertained on lexicon and phonology. Through these features, elements of the dialect's heterogeneous constitution are evident. Exchange of linguistic material has played a considerable role, brought about and shaped through historical migrations, a maritime livelihood and trading activities shared with surrounding peoples, and influences from the nationally prominent language Persian.

References

Dahlgren, S.-O. 2005. "On the Arabic of Arabkhane in Eastern Iran". In *Linguistic Convergence and Areal Diffusion: Case Studies from Iranian, Semitic and Turkic*, ed. by É. Á. Csató, B. Isaksson, and C. Jahani, 161–72. London: Routledge Curzon.

Daniel, E. L. 1986. "Arab Settlements in Iran". In *Encyclopaedia Iranica*, vol. 2, ed. by E. Yarshater, 210–214. London: Routledge & Kegan Paul.

Fischer, W. and O. Jastrow. 1980. *Handbuch der arabischen Dialekte*. Wiesbaden: Harrassowitz.

Grigore, G. 2011. Le système consonantique de l'arabe parlé à Khorramshahr (Iran). *Romano-Arabica* VIII–XI: 93–102.

Holes, C. 2006. "Gulf States". In *Encyclopaedia of Arabic Language and Linguistics*, vol. 2. Leiden: Brill.

Holes, C. 2018. "The Arabic Dialects of the Gulf: Aspects of their Historical and Sociolinguistic Development". In *Arabic Historical Dialectology: Linguistic and Sociolinguistic Approaches*, ed. by C. Holes, 112–147. Oxford: Oxford University Press.

Ingham, B. 1973. "Urban and rural Arabic in Khūzistān". *Bulletin of the School of Oriental and African Studies* 36: 523–553.

Morris, M., J. C. E. Watson, D. Eades and community members. In press. *A Comparative Cultural Glossary of Modern South Arabian*. Oxford: Oxford University Press.

Nadjmabadi, S. R. 2009. "The Arab Presence on the Iranian Coast of the Persian Gulf". In *The Persian Gulf in History*, ed. by L. G. Potter, 129–145. New York: Palgrave Macmillan.

Rosenhouse, J. 2006. "Bedouin Arabic", in *Encyclopaedia of Arabic Language and Linguistics*, ed. by K. Versteegh, Vol. I, 259–69. Leiden and Boston: Brill.

Seeger, U. 2002. "Zwei Texte im Dialekt der Araber von Chorasan". In *„Sprich doch mit deinen Knechten aramäisch, wir verstehen es!" 60 Beiträge zur Semitistik. Festschrift für Otto Jastrow zum 60. Geburtstag*, ed. by W. Arnold and H. Bobzin, 629–646. Wiesbaden: Harrassowitz.

———————. 2009. „Khalaf – Ein arabisches Dorf in Khorasan". In *Philologisches und Historisches zwischen Anatolien und Sokotra: Analecta Semitica in Memoriam Alexander Sima*, ed. by W. Arnold, M. Jursa, W. W. Müller, and S. Procházka, 307–317. Wiesbaden: Harrassowitz.

Seeger, U. 2013. "Zum Verhältnis der zentralasiatischen arabischen Dialekte". In *Nicht nur mit Engelszungen. Beiträge zur semitischen Dialektologie: Festschrift für Werner Arnold zum 60. Geburtstag*, ed. by R. Kuty, U. Seeger and S. Talay, 313–322. Wiesbaden: Harrasowitz.

Shabibi, M. 2006. *Contact-induced grammatical changes in Khuzestani Arabic* (Ph.D. dissertation). Manchester: University of Manchester.

Zarrinkūb, A. 1975. "The Arab Conquest of Iran and its Aftermath". In *The Cambridge History of Iran*, vol. 4, ed. by R. N. Frye, 4:1–56. Cambridge: Cambridge University Press.

On Iranian borrowings in Tocharian

CHAMS BERNARD
Leiden University

It has been recognized since the decipherment of Tocharian that the vocabulary of both Tocharian A and B contain words of Iranian provenance (Tremblay, 2005:422). Only since Hansen's (1940) study has it become clear that the number of Iranian elements in the Tocharian vocabulary is significant. Further studies have been conducted on this particular subject, for instance Winter (1971), Schwartz (1974), Isebaert (1980), and Tremblay (2005). However, these studies lack systematicity and, often the accuracy of the Iranian (and sometimes Tocharian) data is questionable.

The present talk is based on my ongoing research on Iranian borrowings in Tocharian. On the basis of already known, as well as newly established sound correspondences, I will present the results of my research concerning the stratification of Iranian borrowings in Tocharian. Through the establishment of this method it becomes clear which Iranian loanwords where borrowed at which period of time. For instance if Tocharian B has the vowel *e* and Tocharian A the vowel *a* in a word, corresponding to an Iranian *a* vowel, the source language must be an Old Iranian language, as in example 1.

1. Tocharian B *perne* 'rank, glory', Tocharian A *paräṃ* 'id.' is thus a borrowing from OIr. **farnah* 'glory' (pace Hansen 1940:151-52 who suggested it was a borrowing from a Middle Iranian language, either Sogdian or Khotanese).

Establishing the list of all systematic sound correspondences between Iranian languages (in particular Old Iranian) and the Tocharian languages reveals a number of still unrecognized features of Old and Middle Iranian languages in Central Asia.

In this presentation, I will discuss Bactrian borrowings in Tocharian. Tocharian has borrowed words from (1) Proto-Bactrian (ex. 2. and 3.) and (2) from Bactrian (ex. 4. and 5.).

2. Proto-Bactrian **malu* (< PIr. **maðu-*, cf. Bactrian μολο, New Persian *may*) 'wine' → Tocharian B *mālo* 'type of wine' (cf. also Del Tomba 2020:126).

3. Proto-Bactrian **spaxtanīkə* 'servant' (Bactrian σπαχνιιο) → Tocharian B *spaktanīke*, Tocharian A *spaktānik* 'servant' (cf. Schwartz 1974:411).

4. *λαστανγο (/lastang/) 'judgement, verdict' > TB lastāṅk, TA lāstaṅk 'execution bloc' (Schwartz 1974:402-3)

5. σαρλαρο > σαλαρο (/sa:lar/) 'leader' > TB *Salār*, King of Kucha, 5th century (Tremblay 2005:436).

I will also show how modern Iranian languages can help with the etymology of the Tocharian words, as in example 6.

6. Tocharian B *acakarm* /acakə́rm/ 'Hell monster' ← unattested Middle Iranian **aǰa-karm* 'dragon-worm' cf. Sirjāni *kermejek*, Golbāfi *kermejak* 'worm' < 'small (*-ak*) dragon-worm'.

These examples, and a few others, will show that Tocharian borrowed from a number of Iranian languages. What I aim at is to provide a systematic study of all Iranian borrowings in Tocharian, to be able to determine when and from which language Proto-Tocharian and individual Tocharian languages did those borrowings. The study of these borrowings shows that Tocharian had intense contact with a number of Iranian cultures and languages, while it had more limited contact with other Iranian peoples.

With this presentation I hope to demonstrate the relevance of the study of Iranian borrowings to Tocharian in light of the newest discoveries in both language families, and with a systematic method that has not yet been applied to this subject.

References:

Del Tomba, Alessandro. 2020. *The Tocharian Gender System. A Diachronic Study*. PhD thesis: Leiden.

Hansen, Olaf. 1940. "Tocharisch-iranische Beziehungen" ZDMG, 94, pp. 139-164.

Isebaert, Lambert. 1980. *De Indo-Iraanse Bestanddelen in de Tocharische Woordenschat*. Unpublished PhD thesis.

Schwartz, Martin. 1974. "Irano-Tocharica." In *Mémorial Jean Menasce*, ed. Gignoux, Philippe and Tafazzoli, Ahmad, pp. 399-411. Louvain.

Tremblay, Xavier. 2005. "Irano-Tocharica et Tocharo-Iranica" BSOAS 68:3, pp. 421-449.

Winter, Werner. 1971. "Baktrische Lehnwörter im Tocharischen" In *Donum Indogermanicum: Festgabe für Anton Scherer*, pp. 217-223.

Persian of Tehran: towards a new rise of simple verbs?

DORIAN PASTOR[1] (in cooperation with AGNES KORN[2] & CHRISTIAN RAMMER[3])
[1]École Pratique des Hautes Etudes, [2]CNRS, Paris; [3]ZEW, Mannheim

1. Introduction

As stated by TELEGDI (1951) and many other authors, Persian is a language particularly poor in simple verbs. This is compensated by complex predicates formed from a non-verbal element and a light verb, e.g. *harf zadan* "to speak (letter + hit)", *javāb dādan* "to answer (answer + give)", *dorost kardan* "to prepare (right + do)". Some complex predicates have a denominative equivalent with the same meaning (e.g. *motevaje šodan* / *fahmidan* "to understand").

2. New data

BATENI (1990) points out that, contrary to the general tendency, there are several verbs (formed in the last decades) directly derived from substantives that exist besides the (more common) complex predicates; they are accepted by speakers, but not used in the formal language:

Table (1): Some examples of new simple verbs in Persian (BATENI 1990)

tiγidan	*tiγ zadan*	"to defraud"
šalidan	*šal zadan*	"to limp"
šotidan	*šot kardan*	"to shoot"
soridan	*sor xordan*	"to slide"

Data I collected from internet fora, messaging conversations (SMS, WhatsApp) and oral discussions with friends give at hand that additional new simple verbs are emerging in Persian. I found the following verbs:

Table (2): New simple verbs in Persian (my data)

harfidan	*harf zadan*	"to talk"
zangidan	*zang zadan*	"to call"
gušidan	*guš dādan*	"to listen"
fekridan	*fekr kardan*	"to think"
sohbatidan	*sohbat kardan*	"to speak"
čatidan	*čatidan*	"to chat"
aksidan	*aks gereftan*	"to take a picture"
bahsidan	*bahs kardyn*	"to have an argument"
darsidan	*dars xundan* (*xāndan*)	"to study"

(1) *ne-mi-xād* *bā-ham* ***be-harf-e***
NEG-IPFV-want.PRS.3SG together SBJV-talk.PRS-3SG
"She/he doesn't want to talk with me" (source: WEB)

(2) *man* *bāyad* ***be-zang-am*** *be* *mādar=e* *šohar=am?*
I must SBJV-call.PRS-1SG to mother=EZ husband=PC.1SG
"Do I have to call my husband's mother?" (source: WEB)

3. Survey and results

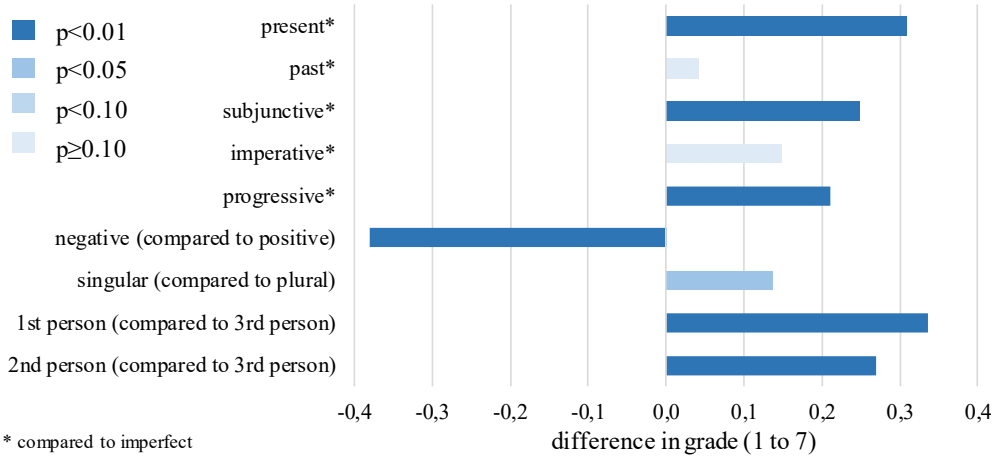
In order to assess the status of the verbs in Table 2, I created an acceptability test (using as models examples such as (1) or (2)) with 45 sentences containing each element of the paradigm (all persons; negative and positive forms; categories present indicative, subjunctive, imperative, past tense, imperfect, present and past progressive); I also included three pseudo-verbs (simple verbs that I made up for the purposes of comparison): **bāvaridan, *yādidan* and **qahridan*. A native Persian speaker who did not participate in the test checked all the

sentences to make sure they were natural enough. Respondents were asked to rate items on a scale from 1 (strongly disapprove) to 7 (strongly approve).[2] The test was answered by 42 respondents between 16 and 58 years of age (31 women, 11 men, all of them from an urban background).

In order to evaluate the influence of the various parameters, I set up a second test, which included all forms of the paradigm for all verbs (i.e. 64 sentences for each verb). Since this questionnaire was very long, only 10 respondents (5 women, 5 men, between 18 and 35 years of age) agreed to participate. For both tests, I read the sentences to the interviewed speakers over face-to-face meetings or meeting in telephone or Whatsapp.

The results (totaling a set of 7010 observations) show statistically significant differences (consistent for both questionnaires) depending on the individual verb, as well as on tense and mood. Interestingly, the 1st and 2nd person are better accepted than the third, and the progressive (the *dāštan* construction) better than the simple present or past. Age plays a significant role as well (with a peak at 35 years on average, but differing depending on the verb). The pseudo-verbs got significantly lower grades and show no age effect.

Figure (1): Acceptance of new denominatives in spoken Persian by tense, mood and person



Results of multivariate regression analysis (controlling for age, gender and verb).
Interpretation: A verb in present tense receives on average a 0.31 better grading than a verb in imperfect.
Light blue indicates statistically insignificant results. These results include both tests.

4. Conclusion

As BUTT (2010: 52) points out, complex predicates are a very productive way to form new predicates since the light verb acts like a verbalizer. An analytic structure has thus developed in Persian, which is part of the general tendency of synthetic structures being marginalized as analytic ones rise. One would expect that this development is unidirectional in the same way as usually assumed for grammaticalisation processes. Thus, the emergence of new simple predicates is unexpected, implying as it does a return to a synthetic structure.

BATENI (1990) observes the emergence of new simple verbs, and my data indicate that the tendency is continuing. It seems that to some extent, the phenomenon is specific to internet fora and the like at present, but it also occurs in everyday speech, and the acceptance rate shows that speakers are not opposed to the new simple verbs.

References

BATENI M. R., « Fārsi zabāni ʿaqim? [Persian, a sterile language?] », in *Adineh* 33, 1368 h.š. [1990], 66-71.

BUTT M., « The Light verb jungle: Still Hacking away », in *Complex Predicates: Cross-linguistic Perspectives on Event Structure*, ed. M. Amberber et al., Cambridge University Press, 2010, 48-78.

TELEGDI Zs., « Nature et fonction des périphrases verbales dites "verbes composes" en persan », in *Acta Orientalia* 2/3, 1951, 315-338.

[2] The approach employed is a Likert rating scale (used to measure attitudes or opinions).

Digor-Ossetic Definite Particle *i*

EMINE ŞAHINGÖZ
Graduate School on Nominal Modification, Goethe-University Frankfurt a.M.

In Digor-Ossetic definiteness can be expressed by the definite particle *i* to emphasize the noun or to prevent ambiguity and misunderstanding. Bailey (1945: 46) reconstructs the particle as Proto-Iranian relative pronoun **ya*-. The definite particle is described by Erschler as "optional and very rare in discourse and written texts" (2019: 9).

(1) Digor

[…]	*'ma*	*ibæl*	*ni-ccijnæ*	*'j*	***i***	***xuarz***	***læg***	*[…]*
	CONJ	CL.3SG.ADES	PV-happiness	COP.3SG.PRS	DEF.PTCL	good	man	

"[…] and **the good man** was happy by this […]" (mægur læg æma wosæ)

Thordarson (2008: 110) defines two possible functions of the "preclitic particle": it either introduces the noun phrase as a kind of definite article. Or it functions as a linking modifier to the preceding head noun as a complementizer.

In Iron-Ossetic this definite particle was lost. With regards to the Iron-Ossetic accentuation rules, where according to Abaev (1964: 12) definiteness can be expressed by an accent movement,[3] Testen states that the "article, which, when counted as the first syllable of its accent complex, caused the accent to be placed one syllable to the left […].When the **i* was lost […] the shifted accent remained grammaticalized as an indication of definiteness" (1997: 729-730). Therefor definiteness in an "accentual complex" could theoretically be expressed by an accent movement onto the preceding syllable, since the lost definite particle *i* would make up the first syllable in this case and thus a new accent arrangement would be necessary. But this can only happen when the second syllable bears the word group accent:

(2a) Iron

syrx	*fæ̲tḳuy*	*myn*	*rad-t*
red	apple	CL.POSS.1SG	give-2SG.IMP.PRS

"give me a red apple"

(2b) Iron

sy̲rx	*fætḳuy*	*myn*	*rad-t*
red	apple	CL.POSS.1SG	give-2SG.IMP.PRS

"give me the red apple" (Thordarson 2008: 110)

The definite particle is not as frequently used as for example the German or English definite article. In the Ossetic Digor Corpus (2.3 million tokens) only 0,08% and in my own corpus (approx. 50 thousand tokens) only 0,6% of the tokens consist of the definite particle *i*. Although rarely used, the particle seems to be obligatory in the positions it is found. After

[3] The accentuation rules in Iron-Ossetic depend on the distribution of strong (*a*, *e*, *i*, *o*, *u*) and weak (*æ*, *y*) vowels; the first two vowels in a word or word group are crucial to decide which syllable is accentuated. If the first vowel is strong, it usually is accentuated. But if the first vowel is weak, the second syllable bears the accent:

1) strong-strong	*xa̲bar* 'news' sg.		2) strong-weak	*bi̲ræ*	'many'
3) weak-strong	*xæzæ̲r* 'house'		4) weak-weak	*fyldæ̲r*	'more'

looking into more than 160 sentences from the Nartic tales[4] that show the definite particle it becomes clear that the particle is not only used to express definiteness but seems to function as a topic-marker at the same time. At exactly half of the 160 sentences it appears in a direct-anaphoric position. This means that the topicalized element was mentioned either in the same sentence or the previous one. Furthermore, the particle in the tales is used to mark commonly known things such as people's designations, discourse new information or it appears in associative-anaphoric position where the topicalized element was not mentioned before but is known due to association. When going through the sentences it becomes clear that the definite particle *i* can mainly be used to pick up a discourse known or presupposed commonly known element after the insertion of at least one element before the topic resumption.

References

ABAEV, Vasilij I. 1949. Ritmika Osetinskoj Reči. In Vasilij I. Abaev (ed.), *Osetinskij jazyk o fol'klor*, 529–560. Moskva-Leningrad: Izdatel'stvo Akademii Nauk SSSR.

ABAEV, Vasilij I. 1964. A Grammatical Sketch of Ossetic (translated by Steven Hill). The American Journal of Linguistics.

BAILEY, H. W. 1945. ASICA. *Transactions of the Philological Society* 44, https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1467-968X.1945.tb00209.x.

ERSCHLER, David (ed.). 2019. *Iron Ossetic* (Handbook of the languages of the Caucasus). Oxford: Oxford University Press.

IVANOV, V. et al., 2001-2018: „Mægwyr læg æmæ us | Mægur læg æma wosæ.“ On: *Dzurut Ironau.* URL: https://ironau.ru/gardanti.html (11.04.2020)

KHAMICAEVA, U, A. & BJAZYROV, A. X. 1989. Narty - Osetinskij geroičeskij épos 2. Moskva: Glavnaja redakcija vostočnoj literatury.

TESTEN, David. 1997. Ossetic Phonology: In: Alan S. Kaye & Peter T. Daniels (eds.), Phonologies of Asia and Africa (Including the Caucasus). 707–731.

THORDARSON, Fridrik. 2009. *Ossetic Grammatical Studies*. Wien: Verlag der Österreichischen Akademie der Wissenschaften.

[4] The Nartic Tales are found among peoples in the northern Caucasus and consists of oral narratives in which the same persons and their relationship to each other is portrayed. The version used for my research is Khamicaeva & Bjazyrov 1989.

Independent outlier: On the linguistic history of Kumzari

ERIK ANONBY
Carleton University, Ottawa, Canada

Kumzari is an endangered language spoken on and around the Musandam Peninsula at the north-eastern tip of Arabia. Although there are fewer than 5000 speakers of the language, it is divided between three countries. In Oman, where most speakers live, it is found in three communities (Kumzar, Khasab and Daba). Kumzari speakers also live in several cities in the United Arab Emirates and in a single village on Larak Island in Iran (Anonby & Yousefian 2011).

Early sources gave conflicting accounts of the origin and linguistic constitution of the language. Likely speaking of Kumzari, Zwemer (1902:57) mentions "a tribe whose speech is neither Persian, Arabic nor Baluchi, but resembles the Himyaritic [South Arabian] dialect of the Mahras". Jayakar (1902:272), the first writer to name and describe the "Kamzàree" language, postulated that the language was "non-Semitic" with "hardly any connection to Arabic", likely originating in the Balochi dialect of "Biyàbool near Minàw [Mināb] on the Persian coast". Miles independently described the language as "a corrupt Persian with a slight admixture of Arabic" (1919:448). In a subsequent sketch of Kumzari, Thomas (1930) states initially that "Kumzāri is largely a compound of Arabic and Persian but distinct from them both" (p. 785), but in a later addendum clarifies that the "grammar and vocabulary show Kumzari to be a quite characteristic Iranian dialect" with origins in "south-central or south-eastern Persia" (p. 843). Referring to certain "established but varying lines of phonetic change", he maintains that Kumzari resembles modern-day "unwritten dialects" of Persia far more than it resembles Modern Persian (p. 844).

In a comparative article on Kumzari along with Lari and Bashkardi, two other Iranic languages in the wider region of the strait of Hormuz, Skjærvø (1989) fleshes out the language's historical connections within the Southwestern (SW) branch of Iranic. Taking Middle Persian as a proto-language, he provides a list of sound changes in the historical evolution of Kumzari. However, because of shortcomings in the quantity and reliability of the available published data (Thomas 1930 and Jayakar 1902), the historical picture of the language is undeveloped. Skjærvø draws numerous areal parallels among Kumzari, Lari and Bashkardi but does not attempt to characterize Kumzari's position within SW or account for the patterning of Arabic structures in the language. Van der Wal Anonby's (2015) grammar of the language affirms the idea of Kumzari as a mixed language with major Persian and Arabian components, along with various other influences. The Arabian component of the language has been treated in depth in van der Wal Anonby (2014) and Anonby (in press). Both studies confirm the Arabian component's considerable time depth, predating the lexification of New Iranic with Arabic.

This paper furthers inquiry into the historical development of Kumzari through detailed re-examination of its Iranic-origin phonological structures in particular. Analysis is based on a full and stable data set, including a 4000-item lexicon gathered during two six-month periods of field research in Khasab and Kumzar and subsequent periods of verification.

Our findings give a fine-grained picture of historical change, and point to three key conclusions: first – perhaps surprisingly, given the distinctive structural patterning of the language – that the Iranic component of the language is situated unambiguously within the Southwestern (SW) branch of that family, and likely even directly descended from Middle

Persian (Set 1); second; that, as an outlier among modern SW Iranic languages, Kumzari has been cut off from Persian during the New Iranic period, exempt from sound changes in the New Persian sphere of influence (Set 2); third, it has undergone a number of independent internal innovations. Some of these are due to Kumzari's Arabian substrate, and subsequent influence from regional Arabic, but all have shaped phonological structures in the lexicon's SW component in significant ways (Set 3).

Selected data (full data set includes 6-10 examples for each sound shift/correspondence)

Set 1: Situation of the Iranic component of Kumzari within Southwestern

1) hardening of MIr *dz** to *d* (K *dānid-*, NP *dānest* 'knew')
2) maintenance of MIr *z** (K *ẓank-*, MP, NP *zan* 'woman')
3) weakening of *g* in MIr *-ag* (MP *xānag*, K *xānaġ*, NP *xāne* 'house')
4) (some SW) weakening of MIr *b* to *w* after low vowels (MP, NP *šab* 'night', K, Lo *šaw*,)
5) (some SW) weakening of postvocalic *d* ("Zagros d") (MP, NP, *dāmād* 'groom', K *dāmar*, Lo *duma*[ð])

Set 2: New Iranic sound changes in which Kumzari has not taken part

1) allophonic shift of *w* to [v] (MP, K *dēw*, NP *dīv* 'ogre')
2) shift of stress to final syllable (MP *'xānag*, K *'xānaġ*, NP *xā'ne* 'house')
3) restructuring of word-initial CC clusters (MP *spēd*, K *spēr*, NP *sefid*, Lo *espid* 'white')
4) (some SW) merging of long mid vowels *ē*/*ō* with long high vowels *ī*/*ū* (MP, K *kē*, NP *ki* 'who?'; MP, K *kōr*, NP *kur* 'blind')
5) (some SW) raising of *ā* before nasals (MP, K *nān*, coll. NP, Lo *nun* 'bread')

Set 3: Kumzari-internal innovations

1) phonologization and lexical diffusion of emphatic consonants (MP *taxl*, NP *talx*, K *ṭalḥ* 'bitter')
2) across-the-board recasting of *z* as *ẓ* (MP, NP *zād*, K *ẓād* 'gave birth')
3) fronting of *u* to *i* closed syllables (MP *gušnag*, NP *gorosne*, K *gišnaġ* 'hungry')
4) reduction of long vowels in unstressed syllables (MP, NP, *haštād*, K *aštad* 'eighty')
5) lengthening of short vowels in stressed syllables (MP, NP *kamar*, K *kāmar* 'waist, lower back')
6) across-the-board shift of *h* to glottal stop (MP *kahwan*, NP *kohne*, K *ka'naġ* 'old')

Abbreviations

K: Kumzari; Lo: Lori; MIr: Middle Iranic; MP: Middle Persian (MacKenzie 1971); NP: New Persian.

References

Anonby, Erik. (in press). Emphatic consonants beyond Arabic: The emergence and proliferation of uvular-pharyngeal emphasis in Kumzari. Linguistics 58:1, https://doi.org/10.1515/ling-2019-0039.

Anonby, Erik and Pakzad Yousefian. 2011. Adaptive multilinguals: A survey of language on Larak Island. Uppsala: Acta Universitatis Upsaliensis.

Jayakar, Atmaram S.G. 1902. The Shahee dialect of Arabic. Journal of the Bombay Branch of the Royal Asiatic Society 21, 246-277.

Miles, Samuel B. 1919/1994. The countries and tribes of the Persian Gulf. Reading, UK: Garnet Publishing.

Skjærvø, Prods Oktor. 1989. Languages of southeast Iran: Lārestānī, Kumzārī, Baškardī. Compendium linguarum iranicarum, ed. Rüdiger Schmitt, pp. 363-369. Wiesbaden: Reichart.

Thomas, Bertram. 1930. The Kumzari dialect of the Shihuh tribe (Musandam), Arabia, and a vocabulary. Journal of the Royal Asiatic Society 62, 785-854.

van der Wal Anonby, Christina. 2014. Traces of Arabian in Kumzari. Proceedings of the Seminar for Arabian Studies 44 (supplement): Languages of Southern Arabia, 137-146.

van der Wal Anonby, Christina. 2015. A grammar of Kumzari, a mixed Perso-Arabian language of Oman (doctoral dissertation). Leiden: Leiden University.

Zwemer, Samuel M. 1902. Three journeys in northern Oman. The geographical journal 19, 54-64.

Grammaticalization of Modals in Persian

FAHIMEH TASALLI BAKHSH
University of Tehran, Tehran, Iran

Grammatical representations of modality in Persian have long been a subject of linguistic research but mainly restricted to studies on moods or on modals such as "*bāyad*" and "*šāyad*". This article considers the evolution of some less-known Persian modals, such as "*tavān(istan)", "residan"* and "*engār*". Here the aim is to study how certain lexical items evolve into grammatical forms to represent modality in Persian. The paths of change are looked at through the lens of grammaticalization focusing on the mechanisms such as reanalysis and analogy and measuring principles such as divergence and de-categorialization (see Hopper and Traugott, 1993).

To avoid prolixity, the evolution of "*tavān(istan)*" is fully discussed here and the grammaticalization process for other modal devices is briefly described:

In Old Persian, some verbs (ex. 1), noun and adjectives are found originated from the root √*tav* meaning "to have the physical ability to do something".

(1) *tyamaiy tanūš tāvayatiy* (DNb, 33-34): inasmuch as my body has the strength… (Kent, 1953: 140).

From the same root, the noun "*tuwān*" is widely used in Middle Persian texts meaning both physical (2) and mental (3) ability to do something.

(2) *čē ān ī kas appurdan nē tuwān* (*Mēnōg ī Xrad*, 39: 7): What is something that no one can steal?

(3) *dānišn ī mardōmīh ayāftan ud dānistan nē tuwān* (*Mēnōg ī Xrad,* 12: 3): [they] cannot obtain and know human knowledge.

"*tuwān*" is specifically used in a repetitive formula with a personal complement (mostly an enclitic pronoun), an infinitive underlying the ability and a copula to imply that "there is (not) ability for someone to do something/ there is (not) ability of (doing) something for someone". This can be automatically interpreted as "someone has/doesn't have the ability to do something":

(4) *ēč kas wizārdan nē tuwān būd* (*Wizārišn ī Čatrang*, 14): Nobody had the ability to explain.

(5) *… ā-t pad dādestānagīh ōy weh tuwān zadan…* (*The Pahlavi Rivāyat*, 4: 1): Then you can better defeat him by [acting] according to law.

In most cases as in (5), the copula is omitted and sometimes "tuwān" is preceded by the negation marker "nē". So, the hearer/reader is facing a verbless sentence and considers the word after negation marker as the verb, since "nē" frequently precedes the verb.

Analogy also plays a significant role here. "tuwān" precedes the (short) infinitive - a behavior similar to such modal verbs as "bāyad" and "šāyad" which had already started off the path of grammaticalization and functioned as auxiliary verbs to represent modality.

The reanalysis goes further when the personal complement is also omitted so that the speaker/writer can generalize the concept of ability and attribute it not to a specific individual or group but to every single person:

(6) *ohrmazd guft kū man mēnōg ī a-griftār ham dast ī man griftan nē tuwān* (*šāyist nē šāyist*, 15: 2): Ohrmazd said that: I'm the intangible spirit. No one can take my hand.

Considering the principle of divergence, we can see that apart from its lexical functions, "*tuwān*" in Middle Persian is used in two different branches; on one hand, by the means of analogy and reanalysis, "*tuwān*" goes through de-categorialization, loses its lexical characteristics as a noun and begins to function as an impersonal modal verb to represent dynamic modality. On the other hand, the ending "*–ist*" is added to "*tuwān*" so that it

functions as a verb which can be fully conjugated. In Middle Persian, there are limited occurrences of this verb, all conjugated as the 3rd person singular in past: "*tuwānist*":

(7) *hāmōyēn ērānšahr abāz ō ēw-xwadāyīh tuwānist āwurdan* (*Kārnāmag ī Ardaxšīr ī Pābagān, 18: 21*): He could bring the whole Iran under [his] single-rulership.

In New Persian, *"tavān"* survives in all three branches; apart from the lexical forms (ex. 8), the impersonal modal verb "*tavān*" functions only in formal writings and is not used in conversations (ex. 9), while "*tavānestan"* obtains a wide range of use in various contexts (ex. 10) and has found its way into Modern New Persian (11):

(8) *hame dar tavān va farmān ī u* (*Kašful asrār*, vol. 1: 14): everything is [surrounded] in his power and order.

(9) *če tavān dānist ke kārhā čun gardad?* (*Persian translation of Tafsir e Tabari*, vol. 2: 346): How can one knows that how things would go on?

(10) *man agar in soxan tavānestami goftan, xod bedin harb nayāmadami* (*Tārix e Bal'ami,* vol.3: 132): if I could say such word, I would never come to this fight.

(11) *nemituni ye kam dige sabr koni:* Can't you wait a little longer?

Changing the word order is another factor that paves the way of grammaticalization for lexical items such as "*tavān*" (see Hopper and Traugott, 1993: 60-64). In many cases, "*tavānestan"* precedes its object, so that the object has to be expressed by a subordinate clause to help the sentence be fully comprehensible. The subordinate clause obviously contains a bold lexical verb which makes the verbal characteristics of "*tavānestan"* grow pale:

(12) *va be qanimat dāšte'and afv čun tavānestand ke be enteqām mašqul šavand*^{verb} (*Tārix Beyhaqi:* 206): and they've availed themselves of forgiveness whenever they had the ability that they keep themselves busy with revenge.

(13) *tunestan xodešun ro be moqe' beresunan*^{verb} *unjā:* They managed to get there on time.

Another example of a grammaticalized modal is the verb "*residan*". In some specific contexts in New Persian, the verb "*residan*" was grammaticalized to represent dynamic, deserving, and permissive modality after undergoing changes such as metaphorical extension:

(14) *ke rā resad ke konad eyb-e dāman-e pākat* (*Divān-e Hāfiz*: 323): Who can carp about your purity?

On the other hand, in recent usages in Modern New Persian, "*residan*" has survived, through a different path, as a means to represent dynamic modality:

(15) *emruz nemiresam inhame mašq ro benevisam:* I can't do all this homework today.

There are also modal adverbs such as *"engār", "guyi"* and *"pendāri"* all grammaticalized from a verbal source to function as epistemic and evidential devices by going through more or less the same way, changing from lexical verbs "*hangārdan*", "*gaub>guftan*", and the verb phrase "*pad ēd dāštan*"

References:

- Bal'ami, M (1999), *Tārixnāme-ye Tabari*. Translation attributed to Bl'ami. ed. Mohammad Roshan, vol. 3. Tehran: Alborz.
- Beyhaqi, A. (2004), *Tārix e Beyhaqi.* ed. Ali Akbar Fayyaz. Tehran: Elm.
- Dhabhar, E.B.N. (1913). Pahlavi Rivayat. Bombay: the Trustees of the Parsi Puncheyet Funds and Properties.
- Hopper, P & Traugott, E. (1993). *Grammaticalization*. 1st edition. Cambridge University Press.
- Kent, R. G. (1953*). Old Persian: Grammar, Texts, Lexicon*. American Oriental Society.
- *Hāfiz*, Sh. M. (2012). *Divān-e Xāje Shamsoddin Mohammad Hāfez-e Shirazi.* ed. by Mohammad Qazvini and Qāsem Qani.Tehran: Zavvār.
- Meybodi, A., R. (1992). *Kašful asrārva oddatul abrār.* ed. Ali Asghar Hekmat. vol. 1. Tehran: AmirKaibr.
- *Translation of tafsir e Tabari* (1977). ed. Habib Yaghmayi. vol. 2. Tehran: Toos.
- Williams, A. V., (1990). The" Pahlavi Rivāyat" Accompanying the" Dādestān Ī Dēnīg". The Royal Danish Academy of sciences and letters.
- http://titus.uni-frankfurt.de/texte/texte2.htm#mpers

Verbal Inflection in Dashtaki

Seyyedeh Fatemeh Musavi
University of Tehran, Iran

Dashtaki is a Fārs dialect that is spoken in a small village in Fārs province of Iran. Dashtak is a high mountainous area (altitude: 2040m) situated at the feet of three tall mountains, which impedes the ability for villagers to communicate with people from other regions. It is known in contemporary linguistic theory that geography shapes language, and in particular, "mountain geography favours isolation" (Auer, Hilpert, Stukenbrock, & Szmrecsanyi, 2013, p. 38), which may explain the conservative and somewhat archaic nature of Dashtaki's verbal inflection paradigm.

This paper aims to introduce Dashtaki as an inflectional variety of Fārs dialects and attempts to make a depiction of the different constructions used in its inflectional paradigms. The abundant literature on Fārs dialects (e.g. Salami, 2004-2014) lacks attention to the Dashtaki variant, and existing work on Dashtaki (see Morgenstierne, 1960; Fallahi, 2008; Hajiyani, 2009; Dorostkar, Qasemi, Fallahi, & Sheibani, 2015) lacks any detailed account of its verbal inflection system. This study aims to fill a gap in this literature by presenting fieldwork data from Dashtak to provide an overview of its verbal inflection system. This speech data collected included poetry verse readings by local poets as recorded by the village's administration board. Daily conversations between older locals were also recorded with the assistance of locals and officials for the purposes of this study. Two native speakers of the dialect (middle-aged, educated locals) assisted in the analysis of the dataset, including their translation of more than 200 verb tokens which are glossed in *Flex*.

The study's preliminary findings indicate that Dashtaki, like many other Fārs dialects, retains ergative marking on verbal constructions. Specifically, the agent of transitive verbs is marked by a personal proclitic, which is typically attested in the past tense. However, Dashtaki's verbal inflection paradigm displays differences to the majority of other Fārs dialects. Consider the perfect marking in Table 1, which uses ergative marking for transitive verbs *boredan and nādan,* as opposed to the majority of the Fārs dialects, in which marking the auxiliary verb using the prefix *ēst-* is the case. With regards to intransitive verbs, periphrastic constructions comprise of the *-ta-ka-* (perfect) infix and the auxiliary *budan* (present), which is the case with Persian too. The difference between the ergative constructions in the past and perfect tense is that the personal proclitics are used in the past tense with the ancient past stem of the verb (or a shortened form of it), while in the perfect, a perfect participle is used. For example, the 3rd sg. of the past for the verb *xardan*: to eat, is *eš xa*: (S)he ate, while the perfect is *eš xarda*: (S)he has eaten. Further, the pluperfect, also uses an ergative construction which is marked with a personal proclitic and a perfect participle and *bi,* i.e. the 3rd person singular of the auxiliary verb *budan*, e.g. the 3rd sg. pluperfect for *xardan*: to eat, is *eš xarda bi*. Other forms examined in this study include imperative, subjunctive, continuous and passive verb inflections. The continuous form displays a serial verb construction identical to that of Persian. Table 1 includes two intransitive verbs (i.e. *budan*: to be and *šodan*: to go) and two transitive verbs (i.e. *nādan*: to put and *boridan*: to cut) to visualise the inflectional features discussed. This paper concludes the inflectional features of the Dashtaki verb and will finds similarities and differences with the other Fārs dialects.

Herewith, I would like to acknowledge the assistance of Samad Moezzi and Mohammad Ahmadi who helped with the data collection and translation.

Bibliography:

Auer, P., Hilpert, M., Stukenbrock, A., & Szmrecsanyi, B. (2013). *Space in Language and Linguistics: Geographical, Interactional and Cognitive Perspectives.* De Gruyter.

Dorostkar, Z., Qasemi, B., Fallahi, A., & Sheibani, P. (2015). Fe'l dar Guyeš-e Daštak-e Fārs. *Adabiyāt va Zaban-hā-ye Mahalli-ye Irān Zamin*, 25-39.

Fallahi, M. H. (2008). Barrasi-ye Avašetāxti va Vājšenāxti-ye Guyeš-e Daštaki. *Motāle'āt-e Irāni*, 159-180.

Hajiyani, F. (2009, Fall). Barrasi-ye Taṭbiqi va Tārixi-ye Sāxt-e Ergative dar Guyeš-hā-ye Ardakāni, Dašti, Daštaki, Kalimian-e Yazd va Lāri. *Teaching Language Skills, 1*(1), 119-138.

Morgenstierne, G. (1960). *Stray notes on Persian dialects.* Norwegian universities Press.

Salami, A. (2004-2014). *Ganjine-ye Guyeš-hā-ye Fārs* (Vols. 1-7). Tehran: Persian Academy.

Verbal Lexeme	Present		Past		Perfect		Pluperfect		Subjunctive		Present Continuous		Imperative	
	1sg	3sg	1sg	3sg	1sg	3sg	1sg	3sg	1sg	3sg	1sg	3sg	2sg	2pl
budan (to be)	*hessem*	*he*	*bodem*	*bi*	*boda'em*	*boda*	*boda bi*	-	*bem*	*bu*	-	-	*ba*	*bit*
šodan (to go)	*mišem*	*mišu*	*eštem*	*raft*	*ešta'em*	*ešta*	*ešta bodem*	*ešta bi*	*bešem*	*bešu*	*darem mišem*	*dārd mišu*	*bešo*	*beši*
nādan (to put)	*minonem*	*minund*	*em nā*	*eš nā*	*em nāda*	*eš nāda*	*em nāda bi*	*eš nāda bi*	*hononem*	*honund*	*darem minonem*	*dārd minund*	*honu(n)*	*hononi*
boredan (to cut)	*miborem*	*miboret*	*em bori*	*eš bori*	*em boreda*	*eš boreda*	*em boreda bi*	*eš boreda bi*	*boborem*	*boboret*	*darem miborem*	*dārd miboret*	*bobor*	*bobori*

Table 1: Dashtaki Verbal Inflection System.

Bashkardi and the languages of Southeast Iran

GERARDO BARBERA
Naples, Italy

This talk will focus on a selection of common and divergent linguistic features characterizing the Bashkardi dialects within the group of so-called languages of Southeast Iran, most of which are spoken in present day Hormozgan province (see P. O. Skjærvø, "Languages of Southeast Iran: Lārestānī, Kumzārī, Baškardī", *Compendium Linguarum Iranicarum*, ed. R. Schmitt, Wiesbaden, 1989, 363-69). This group includes, beside Bashkardi itself, the prestigious – and universally understood within Hormozgan province - Bandari dialect of Bandar Abbas, and, among others, the dialects of Hormoz, Bandar Khamir, Fin, Minab, Rudan. Most of these have been recently termed "Garmsiri" (from the geographical term Garmsir "the warm lands" [here specifically those of Kerman]) by Habib Borjian, who outlined the sharing of many common traits in "a continuum of closely related dialects extending from the Halilrud river valley in the north down to the Strait of Hormuz in the south" (see his article "Kermān. XVI. Languages", *Encyclopaedia Iranica*, 16, New York, 2017, 301-315, esp. 306 ff.).

The languages of Southeast Iran also include, although with a high rate of divergence with respect to the other dialects of this group, Kumzari in the Musandam Peninsula of Oman and Larak island, offshore Bandar Abbas, and the languages of Larestan, an historical region in the Southeast of Fars province. Larestani languages are also spoken in Hormozgan province, particularly in Bastak and Faramarzan, and in Bandar Abbas by the Evazi community originating from Evaz in Fars province. Moreover, Balochi is found in different areas, but predominantly in the region of Jask in the Oman Sea, bordering on Sistan-o Baluchestan province; and Arabic is spoken primarily in the villages of Northwest Hormozgan, especially in Parsian and Bandar Lengeh sub-provinces, but also on the islands, such as Qeshm and Larak.

Bashkardi, in particular, includes two divergent dialect groups that are found throughout Bashagerd, a mountainous territory east of Minab and north of Jask, and in the immediately bordering areas. They are known as Marzi-Gal or Northern Bashkardi and Molki-Gal or Southern Bashkardi, respectively. The materials that will be discussed were collected from 2002 to 2008 in Garu, a small village lying in Minab sub-province outside Bashagerd proper, where a Southern Bashkardi dialect is spoken, or represent Northern Bashkardi as spoken in Sardasht, an important town in Northern Bashagerd. Pioneering work on Bashkardi was done by Ilya Gershevitch following a challenging fieldtrip in 1956. Unfortunately, very little was published with respect to both what the scholar actually collected and what he subsequently was able to re-work in great details.

Nevertheless, although just in few cases do we have language materials of substantial size and publications are still very scarce or with limited, local distribution, the materials at our disposal today provide us with the possibility of studying linguistic relations with more confidence than ever before. Moreover, the "Atlas of the Languages of Iran" Project coordinated by Erik Anonby at Carleton University (www.iranatlas.net) is an ongoing project showing a particular keen interest in the languages of Southeast Iran.

Historically speaking, the majority of languages spoken today in Hormozgan province are Southwestern Iranian, just as, most notably, Persian, Bakhtiari and the Fars dialects, but

with the exception of Balochi, which is Northwestern Iranian (and Arabic, which is, of course, Semitic).

The Southwestern languages have a number of characteristic historic phonetic features that differentiate them from the languages of the so-called Northwestern group: cf., e.g., most typically the presence of the dental sounds *z* and *s* in Northern Bashkardi, Minabi and Persian *ruz*, Bastaki *loz*, and Southern Bashkardi *res* "day", against the palatal sounds *č*, found in Balochi *roč*. However, some important changes from Old to Middle to New Southwestern Iranian distinguish most dialects from each other, isolating Larestani, and from Persian: all words that in Larestani (Bastaki and Faramarzani) and often in Persian begin with *b* from ancient (Proto-Southwestern-Iranian) **w*, appear as *g*(*w*) almost everywhere but as *v* in Southern Bashkardi. For example, Northern Bashkardi and Minabi have *gwak* "frog", Bandari *gak*, Southern Bashkardi has *vak*, but Larestani has *bok* (from **wak*); similarly, Northern Bashkardi shows *gwaik* "bride", Southern Bashkardi *vayex*, Minabi *gowg*, etc., but Larestani shows *bayü*, *beū* (Bastaki *bey*) (from **wayūg*); cf. also Northern Bashkardi, Minabi *gowz* "wasp", Southern Bashkardi *viz*, corresponding to Larestani *bâz*, *biz* (from **wabz*); etc.

On the morpho-syntactic level, some important features will be remarked, too, such as the following:

a) the use of a prefix *a-* (also *at-*) marking indicative mood and durative aspect in the present and imperfect tenses of verbs, where Persian has the prefix *mi-* (from **hami* "always"): e.g. Southern Bashkardi *a-kan-om*, Minabi *a-kon-om*, Persian *mi-kon-am*;
b) the use of a progressive present and past tense construction, consisting of the copula and a "predicate" locution formed by the proclitic prepositions *a-* (Northern Bashkardi, Minabi, Bandari and even Larestani), or *be-* (exclusively Southern Bashkardi) 'in, to' and the infinitive: e.g. Northern Bashkardi, Minabi *a-kerden-om*, Southern Bashkardi *be-kert-in* "I am doing", cf. also Bandari *a-khowt-um*, Bastaki *a-khat-em* "I am sleeping" (Persian, in the same cases, has the constructions *dâr-am mi-kon-am*, lit. "I have/hold I do", or *dâr-am mi-khâb-am* "I'm sleeping");
c) the distinction between intransitive and transitive verb constructions in the past tenses, i.e. the ergative construction. In the ergative construction, the agent, indicating the logical subject, is a clitic pronoun which, in some dialects, is normally attached to the word preceding the verb, or may also follow it. A particularly interesting and apparently isolated case in Southeast Iran is offered by Southern Bashkardi, where endings and clitics can be found side by side in the stem, alternatively as subject and object markers. As expected, clitics occurring after the endings in the present tense stem are objects (in direct and indirect function): e.g. *a-zan-in=et* 'I will beat *you*'. On the other hand, the endings, which are otherwise normally suffixed as subject markers to the present of transitive and intransitive verbs and to the past stems of intransitive verbs, become object indexing markers when suffixed to past transitive stems, following the stems immediately and preceding the clitics. The following table contains a summary of personal marking patterns found at verb stem level in Southern Bashkardi.

Person marking system in the verbal phrase

TENSE *action*	PRESENT *trans*	PRESENT *intr*	PRETERIT *trans*	PRETERIT *intr*	IMPERFECT *trans*
SUBJECT/AGENT MARKERS	endings	endings	clitic pron	endings	clitic pron
OBJECT/OBLIQUE MARKERS	clitic pron	—	endings	—	endings

On Old Persian *tuvam kā haya* "you, whoever…"

JUAN E. BRICEÑO VILLALOBOS
Departamento de Filología Clásica, Universidad Complutense, Madrid, Spain

It has been traditionally admitted that OP *kā*, attested up to six times in the Achaemenid Royal Inscriptions (five times in DB [Darius, Bīsotūn] and once in XPh [Xerxes, Persepolis H]) should be considered a generalizing particle adding an indefinite nuance to the preceding word, in this case, the second person singular of the personal pronoun OP *tuvam* "you" (Smith 1991: 67 fn 37), with a meaning close to " (who)ever". Furthermore, *kā* is seemingly considered to be a reflex of PIE *k^wo-h_1* and, therefore, a cognate of the adverb Gr. πω "yet, at all" (Beekes 2010: 1264, Dunkel 2014: 464; Schmitt 2014: 198, etc), which operates as 'strong' negative polarity item, a type of indefinite pronoun that it is only grammatical in negative contexts.

I will argue in this presentation that the assumption of the instrumental nature of OP *kā* is unsustainable, especially given the negative polarity sensitivity shown by de-instrumental indefinite adverbs among other IE languages and the fact that, as has already been suggested by Kern (1891:47), OP *kā* (*ka-a* /*kā* or *kă*/ in DB and *ka-ă* /*kă*/ in XPh), along with the relative pronoun OP *haya* forms a free-choice (FC) relative indefinite "whoever". FC indefinites are polarity items that, being of a non-episodic nature, are only grammatical in contexts providing multiple alternatives (worlds or situations) (Giannakidou 2001). Also cf. Haspelmath (1997:48). FC indefinites are mainly of two types: they can be either nominal, as in English <u>*Anyone*</u> *can do that*, this type being expressed by OIr.**kaš-cid* > OP *kašci*, Av. *kasciṯ* "everyone", forms that have developed a universal reading derived from its inner FC semantics (unlike Ved. *káś cid* "anyone" that preserves the FC value), or relative as in English <u>*Whoever*</u> *does that, he must do it now* or *You give that to* <u>*whichever*</u> *person you see fit*). Cf. Horn (2000: 71-107).

Moreover, I will study Neo-Babylonian (*mamnu ša*), Aramaic (*mn zy*), and Achaemenid Elamite (*akka*) indefinite relatives attested in the other versions of the Achaemenid Royal Inscriptions in order to assess to what extent they may have prompted the Old Persian employment of **ká-sa-i̯a* as a FC indefinite relative "whoever". Nevertheless, I will argue that what we see in Old Persian is actually an inherited IE structure, *k^wo-(*H*)*i̯ó*-, also found in the indefinite Ved. *káya*- "anyone, everyone" and other IE languages. Finally, I will show that Middle Persian and Armenian interrogative/relative pronouns, *kē* and *o*(*v*) respectively, are derived from indefinite-relative structures, provided that the derivation of MP *kē* from a fossilized genitive **kahya* (Cf. OAv. *kahiia*, YAv. *kahe*) is typologically problematic and that the use of *kē* in Middle Iranian as a relative pronoun might be better explained by its former use as a relative indefinite.

<u>References:</u>

Beekes, R. (2010), *Etymological dictionary of Greek*, vol 1-2, Boston–Leiden: Brill.

Dunkel, G (2014), *Lexicon der indogermanischen Partikeln und Pronominalstämme*. Bn. 2, Heildelberg: Universitätsverlag Winter.

Giannakidou, A. (2001), "The meaning of free-choice", Linguistics and Philosophy, 659-735.

Haspelmath, M. (1997), *Indefinites pronouns*, Oxford: OUP.

Horn, L. R. (2000), "*Any* and (-)*ever*: Free-choice and free-relatives", in A.Z. Wyner (ed.), The Proceedings of the Fifteenth Annual Conference, The University of Haifa 1999, The Israel Association for Theoretical Linguistics.

Kern, H. (1891), "Interrogativa und Indefinita", in W. Caland (ed.), *Zur Syntax der Pronomina im Avesta*, Amsterdam: Verhandelingen der Koninklijke Akademie van Wetenschappen, Afdeeling Letterkunde 20: 2. pp. 47-50.

Schmitt, R. (1991), *Corpus Inscriptionum Iranicarum. Part I, Inscriptions of Ancient Iran. Vol. I, The Old Persian Inscriptions. Texts I, The Bisitun Inscriptions of Darius the Great: Old Persian Text*, London: published on behalf of Corpus Inscriptionum Iranicarum by School of Oriental and African Studies.

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

The reconstruction of Early New Pashto

JULIAN KREIDL
Indiana University, Bloomington (USA)

Due to the works of Morgenstierne on Pashto etymology (NB 2003) and morphology (1942), and of Cheung on phonology (2011), our knowledge about the Old Iranian origin of a large part of the most common vocabulary and inflectional suffixes has been greatly expanded. However, while the starting point (Old Iranian or Proto-Iranian) and the end (contemporary Pashto) are rather well-known, we still don't know much about how the intermediary stages might have looked like. The earliest written records of Pashto are no great help in this respect.

To give some examples: It has been known since the earliest days of Pashto etymological studies that *num* 'name' must be connected with **nā́man-*, *mā́lga* 'salt' with **namádkā-* < **námadkā-*, *owə́* 'seven' with **haftá-*, *gúta, gwə́ta* 'finger' with **angúštā-*, etc. All four words do not show any important variation, neither chronologically, nor throughout the contemporary speech community. Utilizing Pashto proper cannot help us to decide when *-ā́- was colored to -o-, and yielded -u- before a nasal, or when the first syllables in *mā́lga* and *gúta* were dropped.

However, it is still possible to go further back in time. The next step doesn't help to answer all these questions (notably when *-ā́- turned into -o-), but it still tells us a lot of how fast (or slow) Pashto has changed in the last millennium.

I suspect this next step, "Early New Pashto", was spoken about 800 years ago. If we think that the language of the earliest Pashto works (16th century) is not considerably different from today's language, we would probably still need to go back a couple of hundred years to find the common ancestor of both Pashto proper and Waneci. Already by the 16th century, these two dialects must have diverged notably, and I doubt the differences between them were far smaller than they are today. But such differences could have easily evolved within another 400 years.

I decided to use the term "Early New Pashto" because, despite the many changes I am able to point out, the difference between the contemporary language and the reconstructed language of Pashto proper and Waneci is not big enough to consider it a completely different stage of the language.

Referring to the four nouns from above, only 'finger' has a Waneci form different from Pashto proper: *nəgút*. Together with Pashto proper *gúta, gwə́ta*, we are able to reconstruct an Early New Pashto form **ngúta* (f. *-a is deleted in Waneci); it shows that the initial syllable was still present in Pashto at the time Waneci split off. However, it also implies that all other sound changes which turned **anguštā-* to its peculiar Pashto forms were already completed (i.e. the simplification of *št to t).

Likewise, if we look at the other examples, it seems all phonological changes we assume for today's Pashto had already taken place. For instance:
Old *ā had already been raised to *o, and presumably even to *u before nasals, as **nā́man-* yields *num* in all dialects, including Waneci. (Although I don't want to exclude the possibility that we need to reconstruct **nom* for Early New Pashto, and that o in front of nasals was turned into u independently in both branches).

More important are the differences between, for instance, Pashto proper *plār* vs. Waneci *pyār* 'father'. I suggest the two forms go back to Early New Pashto **pδār*. By supposing the existence of *δ – which can develop easily into both l and a palatal glide –, we can also explain the contrast in *šəl, wšəl* vs. *šwi* (< **šwəy* < **wšəy*) 'twenty' < **wšəδ* and *səl* vs. *si* (< **səy*) 'hundred' < **səδ*.

The ending -ay seems to have been contracted in Pashto proper, cf. *zoy* vs. Waneci *zóya* 'son', with Waneci -a usually being cognate to Pashto proper -ay (cf. *stóray – stóra* 'star'); *zoy* could thus go back to **zóyay*.

Apart from phonology, we can also delve into the past as regards morphology. The highly productive plural marker *-ā́n* is obviously a loan from Persian. The genuine Pashto plural suffix *-úna* is today productive with inanimate nouns, whereas *-ə́*, which almost always triggers stem vowel alternations, is not. I will show that *-ə́* was a common, and productive, plural marker still in Early New Pashto. As -ə́ is probably derived from the ancient nominative plural **-āh* of a-stem nouns, we can explain the two common plurals of *ās* 'horse', *āsā́n* and *āsúna* (earlier *asúna*, with ablaut) as later developments which replaced **aspə́* due to morphological leveling (compare *aspá* 'mare', where the p is still preserved). A similar process is going on in *maẓ̌* 'ram', with its plurals *mẓ̌ə* (still used), *mẓ̌úna* and *maẓ̌ā́n*. I will also try to answer the question about the origin of *-úna*; whether it represents a cognate of the likewise nasalized Avestan *-ånŋhō* (which would be remarkable), or whether it is rather from a genitive plural **-ānām*; and what implication that has for the Early New Pashto case system.

All this suggests that while Pashto has definitely changed over the last, say, 800 years, the few phonological sound shifts in the last millennium imply that Pashto must have changed even faster, and more profoundly, than expected in the millennium before, as a Middle Iranian language. Establishing Early New Pashto (or whatever one wants to call it) will thus help to illuminate the even earlier history of Pashto, as a contemporary of Middle Iranian languages like Bactrian, Sogdian and others.

References:

Cheung, Johnny (2011): "Selected Pashto Problems II. Historical Phonology 1: On Vocalism and Etyma", *Iran and the Caucasus 15*, 169-205. Leiden: Brill.

Morgenstierne, Georg (1942): "Archaisms and Innovations in Pashto Morphology", *Norsk Tidsskrift for Sprogvidenskap,* XII, 88-114. Oslo: Universitetsforlaget.

Morgenstierne, Georg (2003): *A New Etymological Vocabulary of Pashto*. Ed. by J. Elfenbein, D.N. MacKenzie, N. Sims-Williams. Wiesbaden: Dr. Ludwig Reicher Verlag.

Placement of the locational goal elements in Kurmanji Kurdish in contact with Turkish and German

KATERYNA IEFREMENKO
University of Potsdam, Germany

Kurdish is characterized as an OV language; however, it is verb-final only to some extent. In fact, in all varieties of Kurdish, there are elements which are placed in a post-verbal position (Haig 2014). These elements always seem to be goals, according to Haig (2014), i.e. they are locational goals of verbs of motion, recipients of verbs of transfer, and addresses of verbs of speech. According to Gündoğdu (2017), the locational goals of verbs of motion are always placed in a post-verbal position across all Kurmanji dialects.

For many centuries, Kurmanji has existed in contact with other languages. In Turkey, in particular in urban areas in the West, Kurmanji is in strong contact with Turkish, where the latter is dominant. Thus, when it comes to heritage speakers (HSs) of Kurmanji in Germany, they often have two heritage languages, Kurmanji and Turkish, in addition to the majority German language.

Both German and Turkish are SOV languages; however, to different degrees (Erdal 1999). Post-verbal position in Turkish appears to be a structure mostly characteristic to informal register. In the formal register, the post-verbal position rarely occurs. Besides, in Standard Turkish, word order is related to discourse and pragmatic information, where topicalized information is in the sentence-initial position, backgrounded information is in the post-verbal position, and new information occurs in the immediate pre-verbal position (Schroeder 1995, Kornfilt 1997). German is also considered to be a SOV language, although this order is only visible in embedded clauses, and not in declarative main clauses or yes/no questions. Besides, arguments, adverbials, and attributes are allowed in the post-verbal field in German (Frey 2019).

In my presentation, I investigate the placement of locational goals of verbs of motion across spoken and written texts in Kurmanji in contact with Turkish and German. In close contact with Turkish, it appears that goals of verbs of motion in Kurmanji can also be placed in a pre-verbal position, while position of goals seems to be rather robust for Kurmanji HSs in Germany. Thus, I ask whether, and under which circumstances, Kurmanji speakers in the West of Turkey tend to place locational goals also in a pre-verbal position, and on the other hand, what influences robustness of goals in HSs of Kurmanji in Germany.

The analysis is based on the spoken and written elicited data produced by 30 adult speakers of northern and southern dialects of Kurmanji Kurdish in Turkey (Ankara) as well as 16 adult heritage speakers of Kurmanji Kurdish and Turkish in Germany (Berlin).

References:

Erdal, M. (1999). Das Nachfeld im Deutschen und im Türkischen. In L. Johanson; J. Rehbein (eds.), *Deutsch und Türkisch im Vergleich*. [Turcologica 39]. Wiesbaden: Harrassowitz, 53-94.

Frey, W. (2019). Unterschiedliches kommt unterschiedlich an den rechten Rand. Presentation at the University of Potsdam. November, 27, 2019.

Gündoğdu, S. (2017). Goal Constituents Across Kurmanji Dialects. Poster presented on the First North American Conference in Iranian Linguistics (NACIL 1).

Haig, G. (2014). VG Word Order in Kurdish and Neo-Aramaic: Typological and Areal Considerations. In: Khan, Geoffrey and Lidia Napiorkowska (eds.), *Neo-aramaic and its linguistic context*, 407-425. New York: Gorgias Press.

Kornfilt, J. (1997). *Turkish [Descriptive Grammars]*. London: Routledge.

Schroeder, C. (1995). Post-predicate arguments in Turkish. *Sprachtypologie und Universalienforschung*, 48: 204-218.

Multielement Structure of Wakhi Stress in Pamir Region

LIUBOV G. SILANTEVA
Moscow State University n.a. M.V. Lomonosov, Moscow, Russia

Wakhi is one of the Pamir languages in the Eastern Iranian branch of Indo-European languages. Native speakers of Wakhi live in Tajikistan, Afghanistan, Pakistan and China. The language does not have unified writing system, literary norm and educational status, but as a koiné language it occupies a dominant position in Wakhan. Most of the Wakhis are multilingual.

Phonetically, Wakhi is characterized by positional quantitative changes in vowel and by the presence of cerebral consonants. There are distinguished six vowel phonemes: *i, ə, a, o, u, ɨ*. The latter has increased variability and the ability to sound like *ü* and *ö*. In addition, *e* appears only in Tajik borrowings. Vowels are characterized by the inherent duration, which was studied by Sokolova, Pakhalina et al. and Steblin-Kamensky. Reliable duration intervals for the Wakhi vowels were computed by Ivanov. Before this research the Wakhi word stress has not been studied experimentally. It was considered expiratory (dynamic) [2], i.e. a stressed syllable marker was supposed to be intensity.

For our instrumental-phonetic study of the Wakhi word stress two experiments were conducted. All the vowels acted as syllable nucleuses in our experiments. Parameters of syllable nucleuses in stressed and unstressed syllables were compared. The phonetic environment of the syllable nucleuses can be both voiced (sonorous) and unvoiced, and consequently vary in terms of vocal cords work. When we search for an acoustic correlate, it is important to have the same set of parameters for each item. That is why the phonetic environment was not regarded, and we only dealt with the syllable nucleuses themselves.

In the first experiment the speech of four native speakers of Wakhi (two men and two women) recorded only in the acoustic channel was examined. Two- and three-compound numerals were taken as a lexical material studied. The segmentation of the speech signal and the acoustic analysis of the recorded realizations were carried out using Praat at the Laboratory of Experimental Phonetics of IAAS MSU. In each syllable, the duration (T), which was an independent parameter, was measured. The parameters of intensity (I) and pitch (F_0) are correlated via subglottal pressure [3]. We suppose that the interaction of these three main parameters depends on the prosodic system of the language and occurs differently in each language. In addition, derivative integrated parameters were examined: F_0-area represents the area of the figure, bounded by the frequency curve and the time axis; intensity area (I-area) represents the area of the figure under the I-curve; volume (V) is a three-dimensional figure, limited by the I-curve, F_0-curve and the center line. According to our measurements, a table was compiled that was processed by the SPSS statistical package, using a linear multidimensional model. Statistical connection of the stress with the syllable nucleuses parameters was checked. In total 120 syllables were examined.

The results showed a highly significant connection between the stress and all the parameters ($p < 0.001$), which is exceptional for the Iranian languages: in other Iranian languages stress is either quantitative or tonic [4]. Only in one isolated (non-Iranian) language – Burushaski – the same situation is found [5]. At this stage it was possible to conclude that the stress in Wakhi was multicomponent: dynamic, tonic and quantitative. To clarify the role of various parameters in Wakhi prosody, Experiment 2 was planned.

In the second experiment, the speech of three other native speakers of Wakhi (two men and one woman) was recorded using hardware and software installation Real-Time EGG in two-channel mode: the microphone signal was recorded in the first channel, and the glottographic signal – in the second one. In the work of Hussain & Mielke, the Pakistani variant of Wakhi was examined using a glottograph. Since we worked with the Tajik Wakhi,

this part of the study as well as Experiment 1 has undoubted novelty. All the channels were processed by Praat speech analysis software. The parameters obtained were calculated by the SPSS statistical package using a multidimensional model. When using a glottograph, frequency of vocal cords' vibrations (F_0) and quotient (Q) – a coefficient of the glottis openness – are recorded. The common approach is that the vertical larynx position (VLP) is measured, which allows to assess the speech gesture's length. The laryngeal movements can be characterized by an average subsonic frequency, which in our case was determined in the range of 7–37 Hz. Two more parameters of syllable nucleus were added: subsonic frequency (Sub) and vertical larynx position (VLP) (see Fig. 2). A total of 88 syllables were analyzed. An additional channel of information – a glottogram – significantly increased the accuracy of segmentation of the speech in comparison with the single-channel recording: when using two-channel recording, the harmonic component and the noise were presented separately. Similarly to the first experiment, the table was processed by SPSS using a multidimensional linear model. The presence of syllable stress was taken as an independent parameter while the remaining eight parameters were considered to be dependent.

Statistical analysis showed that the parameters of duration (T) and pitch frequency (F_0), as well as the integral parameters F_0-area and Volume, are significantly related to the syllable nucleus's stress. Slightly less relevance is found in intensity (I) and I-area ($p = 0.002$). Laryngeal parameters of infrasound frequency (Sub) and vertical larynx position (VLP) appeared insignificant for the prosody ($p = 0.838$ and $p = 0.897$, respectively).

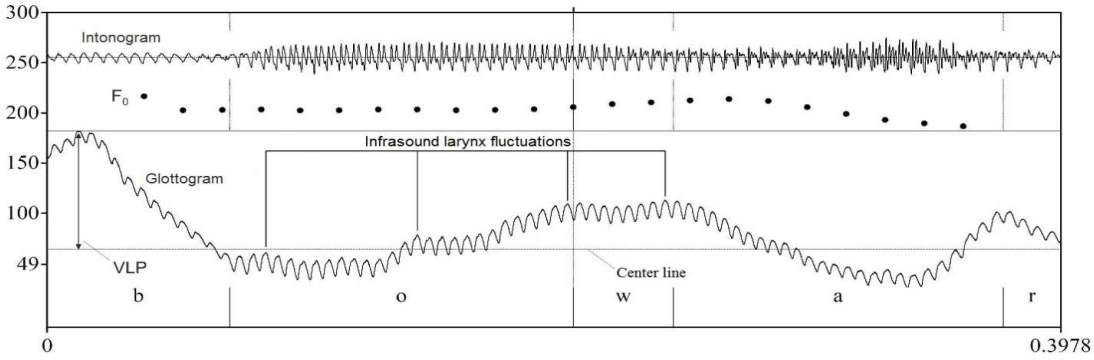


Fig. 2. Segmentation of the word *bowar 'faith'* performed by informant M.: intonogram; F_0-graph; glottogram; infrasound larynx fluctuations; VLP

According to the results of both experiments, the most significant parameters for indicating a stressed syllable appear to be pitch (F_0) and duration (T), which mark stressed syllables in almost all the cases. Intensity (I) and its integral modification (I-area) also rise in the stressed syllable but this occurs less regularly. Statistically significant connection between laryngeal movements and prosody was not discovered at this stage of the study. Thus, the stress in the Wakhi language can be defined as a quantitative-tonic one.

The question of how much the intensity factor is contrasted with other acoustic characteristics of stress and the problem of the connection of laryngeal movements with prosody and intonation requires further analysis.

References

1. Ivanov V.B. Glasnye v rechi vakhancev Pakistana, Tadzhikistana i Kitaya. Moscow, 2019.
2. Pakhalina T.N., Lashkarbekov B.B. Vahanskij yazyk. Yazyki mira. Iranskie yazyki. T. 3. Vostochnoiranskie yazyki. Moscow, 2000.
3. Lehiste I. Suprasegmentals. The M.I.T. Press, Cambridge, Massachusetts and London, England, 1977.
4. Ivanov V.B. Udarenie v iranskih yazykah. Moscow, 2012.
5. Ivanov V.B. Glasnye i mnogokomponentnoe udarenie yazyka burushaski. Moscow, 2013.
6. Hussain Q., Mielke J. Stop laryngeal contrasts of endangered languages of Northern Pakistan. Journal of the Acoustical Society of America (JASA), Vol. 143, #3, 2018.

Areal Typological Study of Contact-Induced Features in Arabic Varieties of Iran

MAHJOUB ZIRAK & FARYAR AKHLAGHI
Research Center for Cultural Heritage and Tourism, Tehran, Iran

Peripheral Arabic dialects manifest various linguistic peculiarities that are mainly due to close linguistic contacts with Indo-European, Turkic and other languages. Historical, geographic, social and other extra linguistic factors have played an important role in the development of Central Asian Arabic dialects and are taken into account in a number of researches conducted on these dialects. Four groups of Central Asian Arabs have been discovered in the region of 1) Bukhara, 2) Qashqa-Darya in Uzbekistan, 3) near Balkh in Afghanistan, and 4) Khorasan Province in Iran (Seeger 2013). There has been extensive research regarding their origin. Most researchers believe that they arrived in early period of Islam with the Arab conquerors to Khorasan (Dahlgren 2005, Jastrow 2014). Ancient Khorasan however had stretched considerably further towards the east and northeast than it does today, encompassing present-day Uzbekistan and Afghanistan. The unsubstantial research conducted to date the languages of these Arabs concludes that their dialect is very strange, with many ancient characteristics, and, since it cannot be categorized under the Arabic dialect types recognized so far, it is to be regarded as Central Asian Arabic (Chikovani 2007). The actual center of Khorasani Arabic is Arabkhane zone (in Nehbandan district) south of Birjand, in South Khorasan. It has also speakers in the district Zir-kuh, around 100 km north-east of Birjand, on the border of Afghanistan and in the outermost north-east of Iran. Some isolated Arabic speaking families live in Sarakhs, on the border of Turkmenistan and in the surrounding areas. The Arab population outer Arabkhane region immigrated to the north parts of Khorasan around 150 years ago.

The Khorasani dialect displays peculiarities that are both similar to and different from those of Standard Arabic and other Arabic dialects. In this article some contact-induced changes resulting from contact with Persian language will be discussed and the degree of these changes among different Arabic varieties spoken in Khorasan province is examined within the framework of areal typology. The material demonstrates the emergence of new linguistic features as a result of long-term contacts with the structurally different language, Persian. Isolation from Arabic speaking Area in centuries contributed to the preservation of several archaic traits (Windfuhr 2005) and, alongside with many traits from the neighboring Persian dialect varieties, the dialect today appears as a strongly contact-influenced variant of Arabic.
Since the Khorasani dialect speakers separated overtime through migration from Arabkhane zone to distant parts mainly Torbat-e Jam, Sarakhs, Dargaz and Kalat, and communication between them has been drastically reduced, their linguistic peculiarities are expected to diverge. This study tries to find more detailed picture of how geographical distance and immigration have impact on dialect diffusion.

In this study we have compared the Arabic dialects of Sarakhs, Dargaz, kalat and Torbat-e Jam with the Arabic dialect of Arabkhane zone as their language origin. We aim to examine how the degree to which the distance competes with areas or regions, aggregates the linguistic distances among language varieties (Mackey 1988). The phenomenon involved here is dialectal “inter-community diffusion” (Boberg 2000). By this term we refer to the observation that, when a linguistic community is separated from its origin by some distance, feature characteristics of one dialect (lexical item, phonetic/phonological, morphological and syntactic linguistic behavior pattern) may turn up in the speech of the speakers of dominant language of the new area.

The corpus of the present study is extracted from Iran Linguistic Atlas (ILA) digital database. ILA funded by the Research Institute for Cultural Heritage and Tourism is a research project for mapping the linguistic diversity of Iran by recording unified data from every spoken dialect of Iran's rural areas by the means of a questionnaire containing words, phrases and sentences. Its initial phase began around the mid-1970s, and the data gathered from Arabic varieties of Khorasan province, villages dates back to 1975.

The Arabic of Arabkhane exhibits not a few features of influence from Persian, whereas at the same time it is easily classified as an Arabic dialect influenced from its neighboring Regional Persian varieties. The rate of innovation in the lexicon, sound changes like the loss of interdentals and loss of "lisping" and changing of the word order confirm accepted views of language change and dialect diffusion through the regional dominant language.
One can therefore conclude that there are more than a dozen Arabic-speaking villages in Khorasan. It seems that Arabic will not remain here long due to strong language dominance from Persian and Arabic gives the impression of a dying language in the region of Sarakhs because of the distance from its origin, Arabkhane.

In the light of the observations and discussions, we will prove that by "the loss of geographically and demographically restricted, or 'marked' variants, and the closely related notion of dialect convergence (O'Shannessy 2011), the minority dialect in the area might become more alike the dominant language through convergent changes. The pattern of geographical diffusion reveals that face-to-face contact of Arabic dialects with the dominant regional Persian will lead to the external influence of the dominant one. Feature characteristics of phonetic, phonological, morphological or syntactic patterns in a minority dialect may come to approximate features of the dominant Language. We suggest that models like the "dominant-minority languages" hierarchy can be refined by a closer engagement with geographical data. This allows an increasingly detailed understanding of language change in relation to patterns of the dialect contact.

References:

Boberg, C. 2000. "Geolinguistic Diffusion and the U.S-Canada Border". In: Language Variation and Change 12, pp. 1-24.

Chokovani, G. 2007. Some Peculiarities of Central Asian Arabic from the Perspective of History of Arabic Language. In: Romano-Arabica VI-VIII (2006-2007): Peripheral Arabic Dialects, pp. 29–38

Dahlgren, S.O. 2005. "On the Arabic of Arabkhane in Eastern Iran." In: Linguistic Convergence and Areal Diffusion: Case Studies from Iranian, Semitic and Turkic. Ed. by Éva Ágnes Csató, Bo Isaksson and Carina Jahani. London, New York: RoutledgeCurzon, pp. 161–171.

Jastrow, o. 2014. "Dialect diferences in Uzbekistan Arabic and their historical implications". In: Alf Lahğa wa Lahğa; Proceedings of the 9th Aida Conference Pescara, Italy 2011 (Neue Beihefte zur Wiener Zeitschrift für die Kunde des Morgenlandes; Bd. 8), pp. 205–212.

Mackey, W.F. 1988. "Geolinguistics: Its Scope and Principles." In: Language in Geographic Context, Edited by C.H. Williams. Clevedon: Multilingual Matters Ltd. pp. 20-46.

O'Shannessy, Carmel. 2011. "Language contact and change in endangered languages." In: The Cambridge Handbook of Endangered Languages. Ed. P. K. Austin and J. Sallabank, Cambridge University Press. pp. 78-99.

Seeger, U. 2013. Zum Verhältnis der zentralasiatischen arabischen Dialekte. In: Nicht nur mit Engelszungen. Beiträge zur semitischen Dialektologie. Festschrift für Werner Arnold zum 60. Geburtstag. Edited by Renaud Kuty, Ulrich Seeger und Shabo Talay. Wiesbaden: Harrassowitz, 2013. pp. 313–322. . (English translation by Sarah Dickins).

Windfuhr, G.L. 2005. "Central Asian Arabic: The Irano-Arabic Dynamics of a New Perfect". In: Linguistic Convergence and Areal Di%usion: Case Studies from Iranian, Semitic and Turkic. Ed. by Éva Ágnes Csató, Bo Isaksson and Carina Jahani. London, New York: RoutledgeCurzon, pp. 111–23.

The emergence of periphrastic aspectual verbs in Persian

MOHAMMAD RASEKH-MAHAND
Bu-Ali Sina University, Hamedan, Iran

In spoken and informal written Persian, instances of two verbs occurring in the same clause sharing the same subject are found frequently. These constructions have some peculiar characteristics. For example, not every verb can be found in these combinations and the verbs that participate comprise a small class, limited to: '*gereftan*' (to take), '*raftan*' (to go), '*āmadan*' (to come), '*nešastan*' (to sit), '*gozāštan*' (to put), and '*bargaštan*' (to come back). The other peculiarity is that these verbs add something to the meaning of the co-occurring verb and if they get omitted, the grammaticality and the truth-conditional meaning of the sentences are not affected. Some examples:

1) a. ***begir*** ***bešin*** *be-bin-am* *či* *mi-g-i.*
Take 3Sg sit 3Sg Subjn-know-1Sg what Prog-say-2Sg
Lit: Take a sit, let me know what you say.

b. ***gereft*** ***xãbid*** *ruye* *taxt.*
Take 3Sg sleep 3Sg on bed
Lit: S/He took a sleep on the bed.

2) *a. nãme rã* ***gereft*** ***pãre*** ***kard.***
Letter ObjM took 3Sg tear do 3Sg
Lit: S/He took the letter and tore it up.

b. ***begir*** *qazã-t* *ro* ***bo-xor****.*
Take 2Sg food-your ObjM Imp-eat 2Sg
Lit: Take and eat your food.

3) a. 'omram ***tamum*** ***šod*** ***raft***
My life finish become went
Lit: My life finished.
b. Ali ***mord raft***
Ali died went
Lit: Ali died.

While this periphrastic verb + verb construction is frequently used in spoken Persian, it is not well explored; only Taleghani (2008), writing on tense, aspect and mood (TAM) of Persian, has mentioned that there are instances of serial verb constructions in Persian. However, she has not referred to this construction.

In this study, I try to explore and explain the different morphosyntactic and functional features of this construction. One of the typological questions to be addressed is whether these multiple verbs construction in Persian are instances of *serial verb constructions* (SVCs). In other words, is Persian a serializing language? While there is not a well-admitted definition of SVCs cross-linguistically, I will try to show that Persian does not have any construction which shows the prototypical features of SVCs. Meanwhile, I will argue that the constructions studied by Taleghani (2008) and grouped as SVCs could not be classified as SVCs, too. Then the next question arises, if these constructions are not SVCs, what are their functions? Could they be classified as auxiliaries and how are they different from proper auxiliaries? How have these structures evolved and what are the motivations for their rise?

I have tried to show that these constructions are emerging as lexical aspectual markers. The first (and sometimes the second) verb in these constructions is acting as an aspectual verb and adds something to the lexical aspect of the main verb. I will try to demonstrate that while these periphrastic constructions are aspectual constructions, they are not yet grammaticalized as proper auxiliaries. I will talk about their syntactic constraints and the way they interact with other items in Persian auxiliary system, and finally, I will try to introduce the functional-typological explanations for the emergence of these periphrastic constructions. The frequency of the minor verbs and their semantics could be the functional motivations behind the grammaticalization of these periphrastic aspectual constructions.

References

Aikhenvald, A. Y. (1999) 'Serial verb constructions and verb compounding: evidence from Tariana (North Arawak)', *Studies in Language 23*.469–508.

Aikhenvald, A. (2006). "Serial verb constructions in typological perspective", In: Aikhennvald, A. Y. And R. M. W. Dixon (eds.), *Serial verb constructions; a cross-linguistic typology*, Oxford University Press, pp:1-68.

Bisang, W. (2009) "Serial verb construction", *Language and Linguistics Compass* 3-3: 792-814.

Bybee, J. (2003) "Mechanisms of change in grammaticization: the role of frequency". In R. Janda & B. Joseph (eds.) *Handbook of Historical Linguistics*. Oxford: Blackwell Publishers.602-623.

Bybee, J. (2007). *Frequency and the Organization of Language*. Oxford: Oxford University Press.

Bybee, J. (2010). *Language, Usage, and Cognition*. Cambridge: Cambridge University Press.

Bybee, J. & Hopper, P. (2001). (eds.), *Frequency and the Emergence of Linguistic Structure*. Amsterdam: John
Benjamins.

Bybee, J. & Thompson, S. (1997). "Three frequency effects in syntax". *Berkeley Linguistics Society* 23. Pp. 378-388.

Jahani, C. (2008) "Expressions of future in Classical and Modern New Persian", In: Karimi, S. et al (eds.) *Aspects of Iranian linguistics*, Newcastle: Cambridge Scholars Publishing, 153-175.

Haspelmath, M. (2016) "The serial verb construction: comparative concept and cross-linguistic generalization". *Language and Linguistics* 17(3): 291-319.

Hopper, P. J. (2008) "Emergent serialization in English: pragmatics and typology". In: J. Good (ed.) *Linguistic universals and language change*, Oxford: Oxford University Press, pp: 253-284.

Taleghani, A. H. (2008) *Modality, Aspect and Negation in Persian*, JohnBenjamins Publishing Company, Amsterdam / Philadelphia.

Phonological Effects in the Ordering of Persian Binomials

MOHSEN MAHDAVI MAZDEH
The University of Arizona

Binomials are coordinated word pairs of the same class (e.g. nouns, adjectives, etc.), e.g. "fine and dandy" or "law and order" in English. A variety of semantic and phonological factors affect the ordering of binomials. Most notably, there is tendency for semantically prominent and unmarked words and phonologically shorter words to come first (Benor and Levy, 2006; Cooper and Ross, 1975; Mollin, 2012).

This study focuses on the role of mora count on the order of binomials in Persian, confirming in passing the role of syllable count too. It is shown that it is the surface form of the binomial, rather than a comparison of the phonological forms of the components in isolation that determines their order. Moreover, this study provides an additional source of evidence for the phonological role of mora count in Persian and the much debated topic of vowel length distinctions in the language.

For this study, binomials were extracted from a corpus of nearly 1.8 million Persian sentences (50m words) obtained from more than 100,000 Persian blogs (Mahdavi, 2012). First, all three-word sequences with the word "o" ("and") in the middle with more than15 occurrences were extracted. The role of semantic factors was minimized by filtering out sequences in which the first word had a higher usage frequency (taking usage frequency an indirect indicator of semantic prominence). Moreover, to focus only on binomials that have a salient "preferred" order, only the pairs that were at least twice as frequent as their reverse were selected. Finally, erroneous items (e.g verb-noun sequences) were manually filtered out. With these filters applied, phonology is expected to be the decisive factor in the ordering of the remaining binomials.

The first clear effect observed was a clear preference for shorter words to come first. Of the 280 pairs that had unequal syllable counts, in 253 (i.e. 90.3%) the shorter word came first. Unlike English but similar to what the literature reports for several other languages, the effect holds in Persian even when the first element has more than one syllable, e.g. *taghdīr-o tashakkor* ("appreciation and thankfulness"), *rādio-o televizion* ("Radio and TV"), *khāhar-obarādar* ("sister and brother").

Rather than the cross-linguistically well-known effect of syllable count, the focus of this study was the effect of mora count on binomials. The first candidate was a naive comparison of the total mora count of the words. For instance, under this metric, the wordbīsh.tar("more"), with one 3-mora syllable and one 2-mora syllable, has a total of 5 moras where as the word *pedar* ("father") has 1+2=3 moras. This naive mora-based comparison returned no significant results; of the 80 pairs in which the two elements had equal syllabic lengths but different moraic lengths, the element with the higher number of moras came first inexactly 40 of them (50%), showing no effect of total mora count whatsoever. Examples of pairs where the moraically longer word comes first include *pīch-o kham* ("twist and curl"), *bīkh-o bon* ("end and root"), and *khesht-o gel* ("adobe and clay"). Examples of the opposite ordering include *chap-o rāst* ("left and right"), *tab-o tāb* ("fever and heat"), and *zar-o zūr* ("gold and power").

Instead of comparing the words in isolation, we may look at the binomial they produce. The interaction of the linking word (for which we assume the pronunciation *o* rather than the

less common alternative *va*) with the first word can have effects on the mora count that can in turn affect ordering. In particular, as shown below, binomials are ordered such that the total mora count of the surface form is minimized if possible.

In most cases, adding *o* to a word does not affect mora count. For instance, *sar* ("head")is 2 moras long but *sar-o* ("head and") is also 2 moras long (two 1-mora syllables) because the coda /r/ now serves as an onset and thus no longer contributes to mora count. Similarly, *dast* and *dast-o* are both 3 moras long. In contrast, a word ending in a long vowel, e.g. *pā* ("foot")has 2 moras, but adding *o* to it results in *pā-o*, which is 3 moras long. Unfortunately, there are relatively few words ending in the long vowels *ā* and *ū*, making any statistical analysis based on them difficult. The long vowel *ī* is a common word-final vowel, but does not show this effect because it loses one of its moras before the vowel *o*.

In spite of this, there are still ways to measure the effect of overall mora count. Consider words ending in a long vowel followed by /n/. As evidenced in Persian metrics, long vowels are generally shortened (lose 1 mora) before a coda /n/. This means that words ending in VVn behave like words ending in VV in terms of their interaction with *o* (adding *o* increases their mora count because the vowel becomes long once the /n/ is no longer a coda) and canthus be used for our study. If the total mora count of the surface form is to be minimized, we expect, for instance, to see *rag-o khun* (4 moras) rather than *khūn-o rag* (5 moras) for "vein and blood". This is indeed the case. This tendency is also responsible for determining the ordering of the names of romantic partners in Persian love stories: *leyli-o majnun* (8 moras) rather than *majnūn-o leylī* (9 moras) and *khosro-wo shīrin* (8 moras) rather than *shīrīn-o khosrow* (9 moras). Phonology seems to take precedence over any potential gender-related factors in these cases as the first name is male in one case and female in the other.

All binomials with frequency above 5 where one word ends in VVn and the other does not change mora count with the addition of *o* were examined. Out of 32 such cases with equal syllable lengths, the word ending in VVn appeared second in 26 of them (81.2%). Of the 6binomials that defied the expected ordering, strong non-phonological motivations (based on semantic prominence, temporal relation, or an influence from the Arabic ordering) could be suggested for the ordering of 3 of them: *ghor'ān-o hadīs* ("The Quran and Hadith"), *īmān-o amal* ("faith and deeds"), and *ghorbān-o ghadīr* (two religious events occurring in the same order in the Islamic calendar).

Finally, it was observed that in many cases this effect was even stronger than the tendency to bring shorter words first. Of the 24 cases where the two tendencies were in conflict, the tendency to minimize overall mora count prevailed in 13 cases (54%).

References

Benor, S. B. and R. Levy (2006). The chicken or the egg? a probabilistic analysis of englishbinomials. Language, 233–278.

Cooper, W. and J. Ross (1975). World order. Parasession on functionalism, 63–111.

Mahdavi, M. (2012). A blog recommender system for persian blogs.

Mollin, S. (2012). Revisiting binomial order in english: Ordering constraints and reversibility. English Language & Linguistics 16(1), 81–103.

Possessor-Source Syncretism in Abşeron Tat

MURAD SULEYMANOV
Institut national des langues et civilisations orientales / UMR 8041, Paris, France

Tat is a group of Iranian dialects closely related to Persian and spoken historically in Azerbaijan and southern Russia. It is divided into two main dialect groups with little mutual intelligibility: Judaeo-Tat (JT) and Muslim Tat (MT), which further branches out into distinct varieties.

The Muslim variety of Abşeron Tat (MT-A) features a peculiar phenomenon whereby the inherited ablative preposition *ä* (*äz* before demonstratives) seems to have extended its original functions (1) to possessive constructions (2).

(1) *qay* *äz* *i* *yazuq* *pul* *xas-tän-i?*
why ABL this poor money want-PRS-2
'Why do you want money from this poor soul?' (MAMMADOVA 2018: 44)

(2) *ru* *äz* *i* *zän* *hičvaxt* *nä-xand-ustän.*
face ABL this woman never NEG-laugh-PRS:3
'This woman's face never smiles.' (MAMMADOVA 2018: 140)

Alongside the construction in (2), there are possessive dependent–head constructions of the opposite word order (with a possessive clitic on the head), which can also occur with the ablative preposition (which remains preposed to the dependent):

(3) *ä* *kilä* *mumsär=i* *uqdä* *duraz* *bu…*
ABL girl hair=POS:3 so_much.DIS long be:PST:3
'The girl's hair grew so long (that)…' (MAMMADOVA 2018: 70)

(4) *äz* *i* *čuqlä* *kilä* *dayi=yi=yum.*
ABL this small.ATR girl uncle=POS:3=COP:1
'I am this little girl's uncle.' (MAMMADOVA 2018: 170)

According to MAMMADOVA (2018: 91), dependents of possessive constructions marked by *ä(z)* are characterised by higher topicality than those in unmarked possessive constructions (e.g. *mumsär kilä* '(a) girl's hair', *ru zän* '(a) woman's face').

The phenomenon of a genitive form developing out of an ablative form (usually through an intermediate partitive function) is typologically attested, especially in Indo-European languages, with Proto-Romance and by extension French (LURAGHI & KITTILÄ 2014: 52) and arguably Russian (DANIEL 2014: 366) being notable examples. Ancient Greek (CONTI & LURAGHI 2014: 444) is known for displaying genitive-ablative syncretism.

A major point of divergence between Tat and other Indo-European languages from the point of view of the genitive-ablative syncretism, or rather the possessor-source syncretism, is the fact that in Tat, a semantic shift towards partitivity on the part of either the ablative or the genitive did not serve as a trigger for such a development. In fact, Tat seems to have well established the ablative preposition as the marker of both source and partitivity, while the head–dependent possessive construction involved neither a genitive preposition, nor a genitive case marker. A cross-dialectal comparison reveals the possessor-source syncretism in MT-A to be the result of a phonetic confusion between the ablative preposition *ä(z)* and what used to be a "possessive particle" that exists in other Tat varieties in the form of *än* (likely related to the Middle Iranian construction *ān ī*, which also marked attributes) and topicalises the possessor (GRJUNBERG 1963: 26), as seen in these Northern MT examples:

(5) *avar-dey* *bä* *xunä* *än* *piyär* *xištän.*
bring-PRF:2/3 LOC home POS:PTCL father self
'…and brought (the sheep) to the house [belonging to] his own father.' (GRJUNBERG 1963: 27)

(6) *zän* *än* *i* *märd=ä* *ʕäyäl=i* *bi-rey.*
woman POS:PTCL this man=OBL child=POS:3 be-PRF:2/3
'The wife of this man [right here] has had a child.' (GRJUNBERG 1963: 27)

The first stage towards the syncretism was probably the loss on the part of *än* of its consonantal element. The instability of word-final *n* is a common trait across Tat varieties (SULEYMANOV 2019: 28), with MT-A **in* > *i* 'this' being one such example. The form *ä* brought forth by this loss (seen in (3)) thus became homophonous with the prenominal variant of the ablative preposition, which in turn led to the possessive *ä* developing the same morphological dichotomy *ä~äz* as the ablative. The process can be represented graphically thus:

ān ī	→	*än*	→	*ä*	→	*ä* *äz*
	loss of ezafe particle		loss of nasal		confusion with ABL *ä*	

The development of the recent constructions in (3–4), where the simple use of the former possessive particle is no longer sufficient to communicate possession and has to be complemented by a possessive clitic, shows that the possessor in them is analysed as a spatial dependent.

The case of MT-A presenting possessor-source syncretism is noteworthy not only because this is the sole Tat variety to show such a phenomenon and not only because its development path is quite individual with regard to its close and distant relatives but also because MT-A is the only language variety to do so in its geographical area, abundant in languages with very elaborate ways of marking semantic roles and spatial relations.

Abbreviations

ABL = ablative, ATR = attributive, COP = copula, DİS = distal, NEG = negative, OBL = oblique, POS = possessive, PRF = perfect, PRS = present, PST = past (preterite), PTCL = particle

Bibliography

CONTI, Luz & LURAGHI, Silvia (2014). "The Ancient Greek partitive genitive in typological perspective". In: LURAGHI & HUUMO, pp. 443–476.

DANIEL, Michael (2014). "The second genitive in Russian". In: LURAGHI & HUUMO, pp. 347–378.

GRJUNBERG, Aleksandr (1963). *Jazyk severoazerbajdžanskix tatov* [The Language of the Tats of Northern Azerbaijan]. Leningrad: Izdatel'stvo Akademii nauk SSSR.

LURAGHI, Silvia & HUUMO, Tuomas, eds. (2014). *Partitive Cases and Related Categories*. Empirical Approaches to Language Typology 54. Berlin / Boston: de Gruyter.

LURAGHI, Silvia & KITTILÄ, Seppo (2014). "Typology and diachrony of partitive case markers". In: LURAGHI & HUUMO, pp. 17–62.

MAMMADOVA, Nayiba (2018). "Éléments de description et documentation du tat de l'Apshéron, langue iranienne d'Azerbaïdjan". Ph.D. thesis defended at the Institut national des langues et civilisations orientales, Paris, https://tel.archives-ouvertes.fr/tel-01801775

SULEYMANOV, Murad (2019). "A Grammar of Şirvan Tat". Ph.D. thesis defended at Université Paris Sciences et Lettres / École Pratique des Hautes Études, Paris.

Differential Object Marking in Iron Ossetic

NATALIA SERDOBOLSKAYA
Institute of Linguistics of the Russian Academy of Sciences, Moscow, Russia /
The Pushkin State Russian Language Institute, Moscow, Russia

Like many Iranian languages, Ossetic exhibits the phenomenon of differential object marking (Bossong 1985): the Direct Object (DO) either takes the genitive marker or remains unmarked, see (1a) and (1b). Abaev (1950: 569–570) and Axvlediani (1969) formulate the rule of the choice of DO marking in terms of both animacy and definiteness: DOs take the genitive if they are animate and definite, and do not take it otherwise. However, the authors state that counterexamples to this rule are widely observed: both genitive inanimate DOs (Abaev 1950: 708; Bagaev 1965: 139) and unmarked animate DOs as in (2). Bagaev (1965) and Gagkaev (1956) attribute the choice of marking to the definiteness of the DO. This study is aimed at testing both factors and presenting a hierarchy of factors influencing the DO encoding in Modern Ossetic (Iron dialect). The source of the data are elicitation sessions with the native speakers.

First, the role of animacy has been tested. Unmarked animate DOs are widely attested in the modern language (2), while inanimate DOs cannot occur with the genitive, and the counterexamples given in (Abaev 1950; Bagaev 1965) are considered by modern native speakers as archaic. My results confirm that the unmarked form is not allowed with pronouns (1st and 2d person; 3d person pronouns have identical forms for nominative and genitive) and proper names (as stated in Abaev 1950). Common nouns denoting people and animals can either take the genitive or remain unmarked. Hence, the role of animacy in DO encoding is illustrated by the following scale:

(i) pronouns > proper names > human-denoting common nouns > animal-denoting common nouns > *inanimates*

genitive — genitive/no marking — *no marking*

Second, I tested the influence of referential properties of the DO on the choice of DO marking. For three of the lexical groups in (i) — pronouns, proper names, common nouns denoting inanimates — no variation inDO marking is observed; referential properties are only relevant for the marking of common nouns denoting humans or animals. For these two groups of nouns, the following results have been obtained through elicitation:

(ii)

1. Definite DOs are always marked.
2. Indefinite specific DOs are divided into two groups, specific known vs. specific unknown in terms of Haspelmath (1997). Specific known DOs take the genitive, while specific unknown do not.
3. Universal and generic DOs can be both genitive-marked and unmarked.
4. Indefinite non-specific DOs are always unmarked.

Third, the lexical class of the noun is also relevant for the choice of DO marking. For example, animate definite DOs must take the genitive, according to the rules (i) and (ii) formulated above. In particular, DOs with possessive clitics take the genitive (3). However, nouns denoting cattle driven in herds can occur unmarked (4) even in this case. By contrast, other animate nouns denoting animals take the genitive.

Thus, the most important factor to predict the DO encoding in Ossetic is animacy; for two lexical noun classes, referential properties are relevant. At that, the principal factor is not definiteness, as argued in reference grammars, rather "familiarity to the speaker".

Examples

(1) a. ***mɜ= mad-ə*** *ragɜj n=al fet:on*
my mother-GEN long NEG=more see-PST.1SG
'I have not seen my mother for a long time.' (Abaev 1950: 569)

b. ***šug*** *ɜr-šɜt:*
wood PV-chop.IMP
'Chop some wood.' (Abaev 1950: 569)

(2) nɜ= qɜd gəc:əl u, ***tɜrquš*** *=dɜr =zə nɜ fen-zənɜ.*
our forest small COP hare PTCL there NEG see-FUT.2SG
'Our forest is very small, there are not even any hares (lit. you will not even see a hare)'.

(3) alan jɜ= ***bɜχ-ə*** */* ****jɜ= bɜχ*** *fɜ-nɜm-ə.*
Alan.NOM his horse-**GEN** his horse PV-beat-PRS.3SG
'Alan beats his horse'.

(4) fəjjaw ***χiž-ə*** ***jæ= quǯə-tɜ*** */* ***quǯə-t-ə.***
shepherd.NOM tend-PRS.3SG his cow-PL cow-PL-GEN
'The shepherd tends his cows'.

References

Abaev, V. I. 1950. Grammatičeskij očerk osetinskogo jazyka. Leningrad,.
Akhvlediani, G. S. (ed). 1969. Grammatika osetinskogo jazyka. T. I. Sintaksis. Ordžonikidze.
Bagaev, N. K. 1965. Sovremennyj osetinskij jazyk. I. Ordžonikidze.
Bossong, G. 1985. Empirische universalienforschung: Differentielle objektmarkierung in den neuiranischen Sprachen. Tübingen: Narr.
Gagkaev, K.E. 1956. Sintaksis osetinskogo jazyka. Ordžonikidze.
Haspelmath, M. 1997. Indefinite pronouns. Oxford University Press.

"Ablative coordination" in Ossetic?

OLEG BELYAEV[1] & IRINA KHOMCHENKOVA[2]
[1]Lomonosov Moscow State University; Institute of Linguistics RAS;
[2]Lomonosov Moscow State University; Vinogradov Russian Language Institute RAS

Overview. In contexts that imply an enumeration of several participants in the same semantic role, Ossetic often uses constructions like (1), where several nouns in ablative singular stand together without an overt conjunct, with the verb having plural morphology:

(1) **foš-ɜj**, **c'iw-ɜj,** **šabi-jɜ** šɜ= fɜllad waχt**-oj**
cattle-ABL bird-ABL child-ABL their labour leave-PST.3PL
'**Cattle**, **birds**, **children** were resting.' (ONC)

At first glance, such examples seem to involve an ablative subject that consists of asyndetically coordinated noun phrases, each of which receives a plural interpretation despite being morphologically singular. Both non-nominative subjects and the plural interpretation of singular NPs are unusual for Ossetic grammar and pose a challenge for grammatical description.

The data. Deeper investigation shows that this impression is false. Ablative nouns can co-occur with an overt nominative plural subject (2), and even with a singular subject that denotes a group and triggers singular agreement (3).

(2) ***žɜrɜd-tɜ*** ***uš-ɜj,*** ***lɜg-ɜj*** *kuv-ənsˑ* …
old-PL woman-ABL man-ABL pray-PRS.3PL
'The **elderly** — **men** (abl.), **women** (abl.) — pray …' (ONC)
(3) *ɜrba-tartˑ-a* *donbetˑər-t-ə* ***rɜʁaw*** *:* ***bɜχ-ɜj,*** ***gal-ɜj***
PV-drive-PST.3SG Dobettyr-PL-GEN flock horse-ABL bull-ABL
'He drove the flock of the Donbettyrs: horses, bulls.' (ONC)

Such examples demonstrate that the ablative NPs do not occupy the subject position.

Furthermore, "ablative coordination" is not, in fact, a syntactically coordinating construction. Unlike ordinary coordination, it does not allow suspended affixation: the ablative affix must be repeated. Nor does it allow the use of overt conjunctions. Thus, neither **uš lɜg-ɜj* (woman man-ABL) nor **uš-ɜj ɜmɜ lɜg-ɜj* (woman-ABL and man-ABL) are grammatical.

Discussion. This construction, then, while being functionally equivalent to coordination, has a different structure: the "conjuncts" are denoted to clause-level adjunct status akin to secondary predication, while another, plural NP that denotes the sum of their referents (either an overt NP or a null subject) occupies the position governed by the verbal predicate. Schematically, this structure can be represented as:

(4) $NP[nom]_{i+j}$ V[pl] $NP[abl]_i$ $NP[abl]_j$

Cross-linguistically, this configuration is highly unusual for coordination-like strategies. Broadly defined, noun phrase conjunction uses one of two strategies: the Coordinate Strategy, where both conjuncts form a constituent occupying the argument position, and the Comitative Strategy, where only one of the adjuncts occupies an argument slot, while the other is a comitative-marked clause-level adjunct (Stassen 2000). Thus either both conjuncts are given equal prominence, or one of the conjuncts is somehow promoted compared to the other.

The Ossetic construction, in contrast, has *both* conjuncts demoted to adjunct status, while the argument position is occupied either by a null pronoun or an NP whose referent is merely a sum of the two conjuncts.

A possible typological counterpart to this construction is the Yakut (Turkic) comitative construction as in (7). In these constructions the presence of a generalizing word or a quantifier is preferred, and the comitative marker can be used bisyndetically.

(7) **ot-tuːn —** **mas-tɯːn** barɯta køɣørøn …
grass-COM tree-COM everything became.green
'**Grass**, **trees**, everything became green …' (Xaritonov 1982)

Examples like (7) are similar to the Ossetic construction, since the referents of comitative phrases are included in the denotation of the generalizing word. There are two main differences: first, according to Arkhipov (2009), this construction might have evolved from a comitative construction [[X-COM] A] meaning 'A including X', but the Ossetic ablative does not have comitative functions and cannot be used in a similar construction. Second, the comitative phrases and the central NP form a constituent, while in Ossetic the central NP and the ablative-marked nouns do not form a constituent.

Another mystery of the Ossetic construction that is unexplained by the above analysis is the fact that ablative NPs always stand in the singular. Note that in all of our examples above, they also lack modifiers — it seems modification is heavily limited, and altogether prohibited in the case of determiners or quantifiers. This suggests that the ablative NPs do not, in fact, denote entities, but are used as depictives and in predicative (or even adverbial) functions, akin to examples like (8), which represent the "functive" meaning as defined in Creissels (2014).

(8) χištɜr-tɜ təng **razi-jɜ** ba-žžad-əštə
elder-PL very glad-ABL PV-remain-PST.3PL
'The (two) elder brothers became very glad.' (Thordarson 2009, p. 162)

Thus, while the use of the ablative in examples like (1) may at first be linked to partitive function (as done, tentatively, in Gagkaev 1956, p. 113), we believe that Thordarson (2009) is correct in treating it as an instance of the predicative function which, ultimately, is related to the instrumental function of the Ossetic ablative.

In the talk, we will examine the features of this construction in more detail, and provide a more detailed syntactic and semantic analysis, drawing on parallels from other languages.

References

Abaev, V. I. (1959). *Grammatičeskij očerk osetinskogo jazyka* [A Grammatical Sketch of Ossetic]. Ordžonikidze. **Arkhipov, A. V. (2009).** "Tipologija komitativnyx konstrukcij." [A typology of comitative constructions] Cand. thesis. Lomonosov Moscow State University. **Bagaev, N. K. (1965).** *Sovremennyj osetinskij jazyk* [Modern Ossetic]. *Fonetika i morfologija* [Phonology and morphology]. Ordžonikidze. **Creissels, D. (2014).** "Functive phrases in typological and diachronic perspective." In: *Studies in Language* 38.3, pp. 605–647. **Gagkaev, K. E. (1956).** *Sintaksis osetinskogo jazyka* [The syntax of Ossetic]. Ordžonikidze. **Stassen, L. (2000).** "AND-languages and WITH-languages." In: *Linguistic Typology* 4.1, pp. 1–54. **Thordarson, F. (2009).** *Ossetic grammatical studies*. Wien. **Xaritonov, L. N., ed. (1982).** *Grammatika sovremennogo jakutskogo literaturnogo jazyka* [A grammat of modern Standard Yakut]. Vol. 1: *Fonetika i morfologija* [Phonology and morphology]. Ed. by L. N. Xaritonov. Moscow.

On One Peculiarity of Lexicalization of Arabic Broken Plural in Persian

PATMAN ANTADZE-MALASHKHIA.
Ivane Javakhishvili Tbilisi State University

It is a well-known fact that in Persian, the Arabic broken plural (BP) forms are sometimes lexicalized in the singular meaning. Such cases are found also in Arabic, and, on their basis, plural forms are produced anew, i.e. double pluralization takes place. „A stronger evidence that BP-s are plurals because of their paradigmatic relations with singulars, and not because they are plurals is that they can themselves feed pluralization“ (Aquaviva 2008:208).

In Persian, there is a stronger tendency of lexicalization based on the transformation of the Arabic BP forms into the singular category. The singular forms are often polysemantic and they coexist with the plural or collective meanings. For instance, ä**sbåb** (singular **säbäb**) 1.„equipment; apparatus;“ 2. „luggage;“ 3. „cause, reason.“ The latter meaning is synonymous with the singular **säbäb**.

The semantic coincidence of the singular and BP forms is found in other examples as well. However, in certain cases, the forms differ as to their stylistic marker. For example, singular **väläd** „offspring; child“ is an archaism, whereas its synonymous BP form **oulåd** is colloquial. The latter also functions as a collective noun meaning „progeny.“

In some cases, the meanings of singular and BP forms differ as to the class of nouns. For instance, **hosn** „beauty“ is an abstract noun, whereas its BP lexicalized form **mähåsen** „beard“ is concrete. This is the case of contamination of the Arabic etymons, because in Arabic **maḥāsin** is the broken plural of **maḥsana** „good, advantage“ and not of **ḥusn**. The word **maḥāsin** itself is lexicalized in Arabic as well. It functions as a collective noun meaning „beauties, charms; attractions; merits; advantages.“ It is this very meaning that motivates Persian **mähåsen** based on the association of **beard** with **beauty.**

The difference between concrete and abstract meanings is also found in the singular **zohr-** „flower“ **(**bookish)[5] and BP **zohur** „glint, radiance“ **(***bookish*). In Arabic, the form **zuḥra** has the meaning of "glint, radiance". Thus, in this case, the meaning of BP in Persian is due to the contamination of the Arabic meaning with the same root and different morphological form.

The difference between concrete and abstract meanings is also obvious if we compare the singular form **qärim** „lender, creditor“ and BP **qorämå** „division of the debtor's property by the creditors“. This example is interesting from the following viewpoint: to a certain extent, the lexicalized meaning of **qorämå** implies plurality, because this meaning refers to more than one creditor.

In Persian, the degree of lexicalization of BP is sometimes so high that BP forms are transformed into adjectives. An example of this is **älvån** 1**"**motley"; 2. „varied“, the singular form of which is **loun** „colour,“. The Arabic etymon is of the same meaning. As for **älvån**, it is found in Arabic with the following genitive in the meaning "all kinds of", which was transmitted into Persian as 1**.**„motley;“ 2. „varied“. These two meanings are also closely associated in other languages. Cf. English - *motley:* 1**.**„disparate;“ 2. „variegated in colour“.

[5] The Arabic etymon **zuḥr** is a collective noun „flowers“.

Thus, in Persian, the relations of lexicalized BP with the corresponding singular form and the Arabic etymons of the same root are diverse. Lexicalization of BP in the singular meaning is chiefly due to the fact that BP forms themselves are the product of stem-forming processes and not of inflection. In addition, as Persian uses suffixes to form the plural, it is more natural for Persian to neglect the function of the forms pluralized without morphemes. However, mostly due to stylistic reasons, Persian frequently uses BP to form the plural and, in some cases, adjusts BP pattern to Persian words proper.

It should also be noted that in certain cases the loan-words borrowed in the plural form produced by means of morphemes are established in the language in the singular meaning. The reason for this is that, for the prevention of redundancy, the plural morpheme of the donor language does not function in those recipient languages which themselves form the plural by means of morphemes. An example of this is the Persian plural form **båno-v-ån** „ladies“ which was transmitted into Georgian in the singular meaning: **banovan-i**[6] „lady.“

Reference

Aquaviva, P. (2008) *Lexical Plurals: A Morphosemantic Approach.* New York: Oxford University Press.

[6] **-i** is a nominative case ending in Georgian.

Towards a historical phonology of Ossetic: relative chronology and Alanic evidence

RONALD I. KIM
Adam Mickiewicz University in Poznań

Although Ossetic is only fragmentarily attested before the 19th century, its historical phonology is generally well understood in comparison with other Modern Iranian languages, partly because of its general conservatism, but also thanks to the efforts of generations of scholars (cf. *inter alia* Miller 1903:11–39, Abaev 1949, Thordarson 1989:459–66, Kambolov 2006). The most important contribution in recent decades is the monograph of Cheung (2002:46– 130), who offers a detailed examination of vocalic developments and consonant changes such as palatalization and gemination. However, with few exceptions, he does not attempt to place the postulated changes into a relative chronology or pursue their consequences for the position of Ossetic within Iranian.

This paper revises and expands the partial relative chronology established in Kim (2003:52–65) in light of Cheung's monograph and subsequent contributions, not least of all the Alanic glosses discovered in an Old Testament lectionary preserved in St. Petersburg (Lubotsky 2015). These glosses, written sometime in the century after 1275, appear to be at approximately the same linguistic stage as the Zelenčuk inscription and the Alanic passage from Tzetzes's *Theogony*, as well as the Jász word list from Hungary: all agree e.g. in preserving POss. **a* before nasals (cf. Oss. *o*) and word-final **-æ* (> D *-æ*, I *-Ø*). The forms ζιρὴν /zærijn/ 'golden' (< **zaranya-*; cf. D *zærijnæ*, I *zærin*) and ἀστέμα̲κ /æstæjmag/ 'eighth' (< **astamyāka-*; cf. D *æstæjmag*) show the diphthongization of **a* > **ai* before sequences of consonant + yod. Furthermore, the apparent apocope in ζιρὴν is better understood as syncope in the compound ζιρὴν κὰ̲μ /zærijn-gam/ 'golden mouth', a literal translation of Χρυσόστομος (Kim 2018:433). Along with syncope, the glosses attest complete loss of word-final **-i* < PIr. **-ah* in κὰ̲μ 'mouth', πά̲ν 'day', στούρ 'great', etc.; its possible retention in τζηρθε 'tombstone' (Zelenčuk; cf. D *cirt*, I *cyrt*) would provide a *terminus post quem*.

Other revisions to Ossetic historical phonology follow from typological considerations. For instance, the change of PIr. **ai*, **au* > POss. **i*, **u* before **n* (e.g. PIr. **wain-* > POss. **win-* > D *win-un*, I *wyn-yn* 'see'; PIr. **gauna-* > POss. **ɣun* > Digor *ɣun*, Iron $q^{w}yn$ 'wool, hair') cannot have proceeded in a single step, as shortening before nasals would be unexpected. Rather, PIr. **ai*, **au* were first diphthongized to **ē*, **ō*, and the latter then became **ī*, **ū* before **n*, which accords with the crosslinguistically frequent raising of vowels before nasals; this stage was reached already by the time of the lectionary glosses, to judge from ζιρὴν 'golden'. Phonemic vowel length was then lost, whereby **ī* and **ū* (including from the few inherited instances of PIr. **ī* and **ū*) merged unconditionally with short **i* and **u*, yielding finally POss. **i*, **u* > D *i*, *u*, I *y*.

The result of these and other revisions is a more precise picture of the phonological evolution of Proto-Iranian through Scythian, Alanic, and Proto-Ossetic to the present-day Digor and Iron dialects. This in turn allows for tentative conclusions on the geographical position of Ossetic in relation to other Iranian languages, above all Sogdian and its modern descendant Yaghnobi. In contrast to early shared features such as the plural in **-tā-* or word-final **-ah* > **-i*, changes such as umlaut, syncope, and perhaps resolution of syllabic sonorants and voicing of voiceless stops spread across a diversifying dialect continuum in the 1st millennium AD, after local innovations had already taken place.

Abbreviations

PIr. = Proto-Iranian; (P)Oss. = (Proto-)Ossetic; D = Digor; I = Iron.

References

Abaev, V. I. [Абаев, В. И.]. 1949. *Осетинский язык и фольклор*. Том I. Москва/Ленинград: Издательство Академии Наук СССР.

Cheung, Johnny. 2002. *Studies in the Historical Development of the Ossetic Vocalism*. Wiesbaden: Reichert.

Kambolov, T. T. [Камболов, Т. Т.]. 2006. *Очерк истории осетинского языка*. Владикавказ: Ир.

Kim, Ronald I. 2003. On the historical phonology of Ossetic: the origin of the oblique case suffix. *Journal of the American Oriental Society* 123.1, 43–71.

—. 2018. Review of Lubotsky 2015. *Folia Orientalia* 55, 431–435.

Lubotsky, Alexander. 2015. *Alanic Marginal Notes in a Greek Liturgical Manuscript*. (Veröffentlichungen zur Iranistik 76. Grammatica Iranica 2.) Wien: Verlag der Österreichischen Akademie der Wissenschaften.

Miller, Wsewolod. 1903. Die Sprache der Osseten. In: Wilhelm Geiger und Ernst Kuhn (eds.), *Grundriß der iranischen Philologie*. Anhang zum ersten Band. Straßburg: Trübner.

Thordarson, Fridrik. 1989. Ossetic. In Rüdiger Schmitt (ed.), *Compendium Linguarum Iranicarum*. Wiesbaden: Reichert, 456–479.

The Features of the "Iranian people with addiction" (IPWA)

SAEID NOROUZI& HASSAN RAFIEY
University of Social Welfare & Rehabilitation Sciences (USWR), Tehran, Iran

Many Scholars have attempted to analyze the slang of people with addiction (PWA) in different countries of the world. Nevertheless, the features of this variety of language have been scarcely treated. The expansion of the slang of PWA is manifested in its use in the everyday speech of non-addicted youth. Some of the slang words and phrases of PWA have been incorporated into the mainstream dialect of society such as [čet], i.e. (Bad trip or disorientation after using Hashish).

The slang of PWA, just like any natural language subsystems, "lives" under the general laws of that languages, and for it's functioning, it adheres well to the general regularities operating in the language.

The main aim of this study is to analyze the features of the spoken language of Iranian people struggling with addiction who consist more than two-thirds of prison population in this country.

Along 22 years of counseling work with PWA and their families, their slang words and phrases were selected, registered, and published as the first book in this field in Iran, i.e. [zabân-e-hâl] (language of the moments) with approximately 2000 entries. The lexicon provides information about headwords, including their definition, examples, and etymology for some entries.

For this purpose, and in line with my counseling work, contextual-dynamic method was used to collect data from the respondents. This method mainly involves the use of observation and dialogue. Then, the entries of the book consisted our data and were subjected to qualitative content analysis.

This study also reveals that the slang of PWA exhibits, figurative language, emotional content, and the use of prison-specific metaphors. The most prevalent processes of word-formation are as follow:

1. Semantic/ metaphorical extension
   [masaleh foruš]: drug dealer; [lačak-i]: Shishe (a type of stimulant drugs); [qir kâr]: addicted to opium; [panir]: Hashish; [xar mast]: heavy drunk; [xortum-i], [damâq-i]: sniffing
2. Neologism
   [nasax]: withdrawal from stimulant drugs; [gabr-i]: mixing of opium and residue opium
3. Subtraction
   [lul-e-lul]: intoxicated; [kerkere]: potent heroin
4. Shortening
   [ânti]: Antihistamine; [ef]: Pseudoephedrine; [alk]: Alcohol; [nor]: Norgesic (A type of drug for detoxification)
5. Borrowing
   [peti]: sniffing (Lori dialect); [melo]: mild; [angebâ]: Hashish (Motrebi dialect); [over]: overdose; [igâresâ]: cigarette (Motrebi dialect); [dax]: good, high quality (Gypsy dialect); [nâdax]: bad, low quality (Gypsy dialect)

6. Clipping
   [tal]: talxaki, (opium); [qaran], qarantine (quarantine); [tez]: tazriq (injection); [doki]: Doctor
7. Blending
   [mas čet]: mast čet, bad trip and drunkenness; [qol fur]: qol qol-i and Vaphor (a handmade paraphernalia)
8. echo reduplication
   [tal-o-tul]: opium; [ti ti]: opium
9. Onomatopoesis:
   [dʒir dʒir]: (sound of opium smoking by Vaphor); [qol qol-i]: (sound of smoking by a handmade paraphernalia)

Among the slang words and phrases of IPWA, similarities with a few English utterances can be seen in various forms, ranging from [goriz miše ye riz] (one drink – one drunk), [šaqâyeq] (Bloodshot eyes), [qofl-i] (stoned), and so on. In a nutshell, it seems that the slang of PWA is a highly creative one.

References:

Eble, C. (1996). *Slang and Sociability*. London and Chapel Hill: University of North Carolina Press

Einat, T. & Einat, H. (2000). *Inmate argot as an expression of prison subculture: The Israeli case*. [in:] The Prison Journal. Vol. 80, No. 3, pp. 309-325.

Harris, B. (2012). *What Is Argot?* (n.d.). Retrieved November 20, 2012, from wise GEEK:http://www.wisegeek.com/what-isargot.htm. Conjecture Corporation

Hui, W. (2009). *A Brief Analysis of the Characteristics of crime Argot.* Unpublished theses, Guangdong Police College; Guangzhou, China.

Norouzi, Saeid (2015). *The slangs of addiction (zabane – hall).* Danjeh publication.

Ogutu Peter Okoth, Opande Nilson Isaac, Oluoch Stephen. Kiswahili ya Jela (2018). *The Features of the Kenyan Prison Argot*. International Journal of Language and Linguistics. Vol. 6, No. 6, pp. 185-196.

Verbal particles in Laki: Patterns of Polysemy and Development of an Elusive Lexical Class

SARA BELELLI
Postdoctoral researcher, "L'Orientale" University of Naples / ISMEO, Italy

Verbal particles – often called 'preverbs', despite their varying placement within the Verbal Phrase – are most often treated in grammatical descriptions of Iranian languages in the context of verb compounding. Indeed, this closed class of elements standing in between lexical and function words is involved in the formation of a particular category of complex predicates, by and large comparable to Eng. *phrasal verbs*, Ger. *Partikelverben*, It. *verbi sintagmatici*, Fr. *verbes à preverbe*, etc. Another well-known feature of verbal particles – common to many IE languages – is that they occur in equal or very similar shape also with Noun Phrases, as adpositional elements introducing various arguments and semantic roles.

This contribution attempts at investigating function, meaning and origin of some verbal particles typical of Laki, a still poorly-known cluster of Iranian varieties spoken at the southern periphery of the Kurdish language continuum. The data at our disposal allow for the identification of at least four such items in the target language (cf. Lazard 1992: 221-222, Fattah 2000: 433ff.):

1) *ö(w)r* 'up'
řün-á ma-gir-ē=ör=ī
butter-DEF IND-take.PRS-3sg=PART=3sg
'The butter, he picks it up/puts it aside' [Laki-Kermānshāhi, Harsin, Belelli 2016, 5:28]

2) *ēr* 'down, in/out'
das=a ma-n-īm=ē=ēr
hand=IND IND-put.PRS-1pl=3sg=PART
'Nous lui mettons les bras dédans' [Laki of Aleshtar, Lazard 1992, IV.7]

3) *(i)rā* 'forth, around'
māł-a gird-im=ē=rā
house-DEF take.PST-1sg=3sg=PART
'La maison, je l'ai assiégée; la maison, je l'ai cernée [Laki Kermānshāhi, Harsin, Fattah, 2000:448]

4) *ā* 'back, re-'
hāt-im=as=ā
come.PST-1sg=COP.PRS.3sg=PART
'I've come back' [Laki Kermānshāhi, Harsin, Belelli 2016, 6:113]

These forms – alongside possible postpositional equivalents involved in circumpositions – seem specific to Laki and neighbouring Laki-Kermānshāhi varieties, and are therefore included among the dialectologically significant traits distinguishing them from adjacent languages and dialect clusters (esp. the bulk of Southern Kurdish dialects, cf. Fattah 2000: 13, 61-62).

The investigation will review some occurrences of these particles in published texts, lexicographical sources, and collected speech samples. Comparisons with similar items in other Iranian languages will allow for the elaboration of preliminary hypotheses on the patterns of their historical and semantic development.

Selected bibliography:

Belelli, S. 2016. *A study on Language and Folklore in the City of Harsin (Kermānshāh Province, West Iran): Sketch Grammar with Texts and Lexicon*, PhD thesis, "L'Orientale" University of Naples.

— (forthcoming) "Una Leggenda in laki da Darb-e Gonbad (Lorestān, Iran)", *AION-or*.

Bubenik, V. 2006. "Cases and Prepositions in Iranian". In John Hewson & Vit Bubenik (eds.), *From case to adposition: The development of configurational syntax in Indo-European languages*, pp. 131-159.

Fattah, I. K. 2000. *Les dialectes kurdes méridionaux: étude linguistique et dialectologique*, Acta Iranica 37, Leuven: Peeters.

Jazab, Y. K. & L. Rzehak. "Deictoids /rā-/, /dər-/ or /dar-/, and /wər-/ or /war-/ of some Karlanri Varieties", *Pashto* 42/645-646, pp. 45-53.

Kiyāni Kulivand, K. 1390/2011. *Farhang-e kiyān: farhang va vāženāme-ye laki* [Kiyān Dictionary: Laki Dictionary and Glossary], 2 vol., Khorramābād: Sifā.

Lazard, G. 1992. "Le dialecte laki d'Aleshtar (kurde méridional)", *Studia Iranica* 21, pp. 215-245.

— 1995. "Préverbes et typologie", in André Rousseau (ed.), *Les préverbes dans les langues d'Europe: Introduction à l'analyse de la préverbation*. Villeneuve d'Asq: Presses Universitaires du Septentrion, pp. 23–30.

Moshkalo, V. V. 1988. "The reflections of the Old Iranian preverbs on the Baluchi verbs", *Newsletter of Baluchistan Studies* 5, pp. 71-74.

Stilo, D. 2009. "Circumpositions as an Areal Response: The Case Study of the Iranian Zone", in Lars Johanson (ed.), *Turkic Languages* 13, pp. 3-33.

Telegdi, Z. 1955. "Beiträge zur historischen Grammatik des Neupersischen. 1. Über die Partikelkomposition in Neupersischen", in *Acta Linguistica Academiae Scientiarum Hungaricae* 5/1-2, pp. 67-183.

Thordarson, F. 2011. "Preverbs in Ossetic: Some Reflections", *Nartamongæ* VIII/1- 2, pp. 219-228.

On the Developments of *Bāyestan*: How *Bāyad* and *Bāyest* Lost Their Verbal Traits

SETAYESH DASHTI
University of Tehran, Tehran, Iran

Bāyestan and its cognate forms have been subject of debate among grammarians of Persian language. Its two main forms *bāyad* and *bāyest* and their variants still remain in the contemporary Persian language, but its functions and forms have undergone many changes from Early New Persian until now. In the early stages of the Persian language, it was used as an ordinary verb with the complete conjugational system. Today, grammarians describe it as a modal verb, auxiliary verb or adverb (e.g. Bateni 2005: 123, Lazard 2006: 130, Labbafan Khosh & Darzi 2015: 97-112). Whatever we label it in the Persian language of 20th and 21st century, we still have to deal with a process in which a fully conjugated verb has lost a number of its verbal traits. The purpose of this research is to analyze this process and its motives. To prove the facts given about the frequency of grammatical structures in this research, I have studied approximately the first 20 pages of 10 prose books from each century of Persian literature history (from 10th century onwards). I have also investigated some data from the corpus of the Academy of Persian Language and Literature. Examples here are also taken from texts written in Early New Persian and Classical Persian.

In Early New Persian, *bāyestan* had 3 different forms:

1. A conjugated verb with two stems *bāy-* and *bāyest-* that received personal endings, which could be used either with an indirect object (which normally comes before the subject): بازرگانی را دو تن بایند: خرنده و فروشنده or with a predicate: گرفتمت که شدی آن‌چنان‌که می‌بایی;
2. A verb whose subject is an apocopated or full infinitive (and thus is always conjugated in 3rd person singular). This form can be personalized with an indirect object with the marker *rā*: ما را پیش باید رفتن که هم بیمِ مرگ است و هم امیدِ راحت;
3. A verb whose subject is a subject clause. The subject of this subject clause can also be realized as the indirect object of *bāyestan*: باید که تو به هندوستان خلیفتی شایسته بنشانی and مرا همی‌باید که ترا ببینم.

In Modern Persian, **the first form** does not exist, except for its predicative construction only for 3rd person singular: غذا اینطور باید. **The second form** is only used with an apocopated infinitive, never personalized. **The third form** is still productive, but never with an indirect object of *bāyestan*.

Based on my observations of its usage in Persian texts, there were supposedly three main steps in this process and a fourth one, which is rather an outcome of this process. The first three ones are as follows:

1. *bāyestan* ceases to be conjugated in all persons and is confined to the third person singular: *bāyad*, *bāyest*;
2. In forms where the subject of *bāyad* /*bāyest* is a subject clause, two forms coexist. One where the subject of the subject clause is demonstrated by the indirect object of *bāyad*/*bāyest* and comes before *bāyad*/*bāyest* with the indirect object marker *rā*, the other where *bāyad*/*bāyest* has no indirect object, but the subject of the subject clause is topicalized and shifted before the main verb, *bāyad*/*bāyest*;
3. In forms where the subject of *bāyad*/*bāyest* is an infinitive, it becomes no longer personalized by an indirect object. The fourth one which is much later, is that
4. *bāyad* is no longer restricted to present usage and *bāyest* to past usage, and they are used interchangeably with verbs in past and present.

In what follows, I will explain each of these steps, their grounds and their contributions to the development of *bāyad*/*bāyest*.

In the first step, the conjugation of *bāyestan* becomes restricted to the third person singular. The reason for this development is the substantial difference between the frequency of its

conjugation in the third person singular and other forms. Taking into account that in many cases of its occurrence its subject is an infinitive or a clause, this conjugated form in 3rd person singular becomes so widespread that the speakers do not tend to conjugate it anymore. This tendency is proven by the fact that the latest instances of its conjugation are no later than the 14th century, and very few after the 12th century.

Meanwhile, another feature is arising. Normally, the sentence with a subject clause as the subject of *bāyestan* must have been formulated this way: مرا باید که بروم or باید که من بروم. However, there is a plentitude of examples in Persian texts from the 10th and 11th century onwards where this structure is formulated so: من باید که بروم. Since the omission of the indirect object marker *rā* is not common, we have to analyze these forms as topicalization of the subject of the subject clause. Coexistence of both forms, one with the indirect object of *bāyestan* and the other with the topicalized subject of the subject clause in one text testifies to my explanation as well.

After these two developments, the speaker starts to forget that *bāyestan* was once conjugable and gradually *bāyad* and *bāyest* become frozen words in the language. To this contributes another process: the relative pronoun *ke* tends to become omitted between subordinate clauses and the number of parataxis grows in the language (Lazard 1963: 488). At this point, when the speaker encounters forms such as من باید بروم, he/she recognizes a conjugated verb and a noun/pronoun with which the verb is in accordance. Therefore, the speaker perceives the sentence as a simple sentence. Another tendency, the allocation of *rā* for only marking the direct object, gradually makes forms like مرا باید (که) بروم disappear from the language. Why did not the speaker rephrase this sentence as برای من باید (که) بروم (like other cases of indirect objects marked with *rā* in the Persian language)? The answer is that the speaker no longer understands *bāyad*/ *bāyest* as a verb that can govern an indirect object. For this reason, forms with similar meaning are in use with an indirect object: برای من لازم است / برای من بایسته است -who are predicative-, but not *برای من باید بروم.

Concerning the forms with an infinitive as the subject, the aforementioned developments (the specification of the use of *rā* for indirect object marker and the fact that *bāyad*/*bāyest* are no longer perceived as conjugated verbs) efface forms like مرا باید رفتن. However, forms with an apocopated infinitive remain as a frozen impersonal construction. The reason why this construction is always with apocopated infinitive and never with full infinitive is that the speaker's grammatical intuition demands a verb in the sentences and since *bāyad*/ *bāyest* has lost its verbal value, the speaker uses apocopated infinitive which, in form, is not different from the third person singular form in simple past.

The logical outcome of all these changes would be that the speaker does not think of *bāyad*/*bāyest* as verbs conjugated in certain tense and person, and would use *bāyad* in with past verbs and *bāyest* in present verbs.

References

Bateni, M. (2005). *Tosif-e Sākhteman-e Dasturi-ye Zabān-e Fārsi*. Tehran: Amirkabir.

Labbafan Khosh, Z. & A. Darzi (2015). "Negāhi Be Jāygāhe Nahvi-e Bāyad va Shāyad". *Journal of Language Researches* 5:2. pp. 97-112.

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

Lazard, G. (2006). *Grammaire du persan contemporain*. Tehran : Institut Français De Recherche En Iran.

The Proto-Zaza-Gorani case system: evidence for convergence and expansion

SHUAN OSMAN KARIM
The Ohio State University; Columbus, Ohio USA

The Zaza-Gorani languages are an understudied sub-branch of the Iranian branch of Indo-European spoken in pockets across the Kurdish-speaking zone. These languages are often considered to be archaizing in character (Paul, 1998a). This description is based on their tendency to preserve case-, number-, and gender-marking (lost in many Iranian languages). Additionally, some Zaza-Gorani languages have animacy distinctions which interact with case-, number-, and gender-marking in different ways (Zazaki (Paul, 1998b), Shabaki (MacKenzie, 1956), etc.). The Zazaki nominal paradigm in particular has a large paradigm, maximally distinguishing 144 paradigm cells (Karim, 2019). This complexity in the Zazaki system is rooted in the interaction of not just case, number, and gender, but also definiteness, animacy and ezafe marking (type of modifier).

This study proposes and argues for a nominal system for Proto-Zaza-Gorani with 2 cases, 2 numbers, and 2 genders, and then traces that system's development into the subsequent daughter languages. This assertion is not particularly controversial, given the systems of other Iranian languages. However, the idiosyncrasies of Zaza-Gorani morpho-syntax, and Hewramî and Zazaki in particular, show rapid expansion from a core Proto-Zaza-Gorani nominal system. In this study, I have examined evidence from Zazaki (Paul, 1998b), Hewramî (MacKenzie, 1966), Zerdeyane (Mahmoudveysi, Bailey, Gorani, Jügel, & Jügel, 2016), Gewrecuwî (Mahmoudveysi, Bailey, Paul, & Haig, 2012), and Shabaki (MacKenzie, 1956). The systems of each of these languages have retained relics of the original system albeit embedded within copious independent (not shared) innovations.

The system of Proto-Zaza-Gorani has structural similarities with Hewramî, although it differs crucially along lines explained by convergence with other KZ languages.

	M	F	M(fam)	F(fam)
DIR.SG	∅	-ə	∅	-a
PREP.SG	-i	-e	-r	-ar
DIR.PL	-i	-e	-r	-ar
PREP.PL	-an	-an	-ran	-aran

Table 1. The Proto-Zaza-Gorani case system

	M	F
DIR.SG	∅	-ə
OBL.SG	-i	-e
DIR.PL	-e	-e
OBL.PL	-an	-an

Table 2. The Hewramî case system

Three key ways in which the Proto-Zaza-Gorani nominal system differs from Hewramî are in the existence of separate direct plural endings for masculine and feminine nouns, in the declension of familial nouns (replaced by borrowings from Central Kurdish in Hewramî) and in the environments which condition the oblique (expanded in Zazaki). In Hewramî, the oblique[7] case is used for complements of prepositions, genitival possessors, adverbials, and emphatic subjects of past transitive clauses (Mahmoudveysi, 2019; (Rasekh-Mahand & Naghshbandi, 2013)). This idiosyncratic set of environments which condition the oblique in Hewramî surfaces as a relic within the larger Zazaki system. One reason Zazaki's case system has expanded to its current size is the fact that the ezafe and case marker have univerbated forming a complex suffix (Karim, 2019). When a Zazaki noun occurs in an

[7] The alternation between direct and oblique case is standard in discussion of modern Iranian languages. This designation does not, however accurately capture the diversity of use for the two cases across Iranian languages.

unmodified form, It employs a case system recognizable from neighboring Kurmancî.[8] The oblique is used for complements of prepositions, genitival possessors, adverbials, direct objects of present tense transitive clauses, and all subjects of past transitive clauses (i.e. not just when emphatic). When a modifier, either an attributive adjective of a genitival possessor, is added, an ezafe-case marker is employed which differs from the unmodified case markers. However, in the subset of those oblique environments which are common to both Hewramî and Zazaki, the complex case marker-ezafe is augmented by a /d/. Zazaki has preserved the Hewramî like Proto-Zaza-Gorani system at the core of its own nominal declension.

I present here not only an account of the Proto-Zaza-Gorani nominal system based on comparison between the modern Zaza-Gorani languages but also a description of the ways that that system has developed into the modern languages. The development of each of the daughter languages is understandable when compared to regional contact languages (primarily Kurdish). Thus, in this paper, I answer specific questions about Proto-Zaza-Gorani and thereby provide a basis for a future comprehensive description of the evolution of these languages.

References

Karim, S. O. (2019). Competition between formatives and the diversity of ezafat. *Presented at the 24th International Conference on Historical Linguistics (ICHL24)*.

MacKenzie, D. N. (1956). Bajalani. *Bullschoorieafri Bulletin of the School of Oriental and African Studies, University of London*, *18*(3), 418–435.

MacKenzie, D. N. (1966). *The dialect of Awroman (Hawraman-i Luhon) : Grammatical sketch, texts, and vocabulary*. Kobenhavn: Kommissionaer : Munksgaard.

Mahmoudveysi, Parvin., Bailey, Denise., Paul, Ludwig., & Haig, Geoffrey. (2012). *The Gorani Language of Gawraju, a village of West Iran : texts, grammar, and lexicon*.

Mahmoudveysi, P., Bailey, D., Gorani, T., Jügel, T., & Jügel, T. (2016). *Parvin Mahmoudveysi , Denise Bailey . The Gorani language of Zarda , a village of West Iran . 34*.

Paul, L. (1998a). The position of Zazaki among West Iranian languages. *Proceedings of the Third European Conference of Iranian Studies; Held in Cambridge, 11th to 15th, 1995*, 163–177.

Paul, Ludwig. (1998b). *Zazaki : Grammatik und Versuch einer Dialektologie*. Wiesbaden: L. Reichert.

Rasekh-Mahand, M., & Naghshbandi, Z. (2013). Argument Strength and Differential Case Marking: Some Evidence from Hawrami. *Western Iranian Languages and Dialects*, *1*(1), 1–20.

[8] Kurmancî, aka Northern Kurdish, is Zazaki's main contact language. It has converged with it in many ways reflected in its nominal. pronominal and verbal forms.

On the phonetic differences between the intonation of Persian statements and Questions

VAHID SADEGHI
Imam Khomeini International University

Introduction

As is well known, languages employ different prosodic strategies to distinguish yes/no questions and statements. A cross-linguistic common intonational cue to questions and statements is the direction of F0 movement at the edge of the phrase. While questions are often associated with a sharp final F0 rise, statements are realized with a final F0 fall (Pierrehumbert and Hirshberg, 1990).
However, many languages use differences in the alignment and scaling of the nuclear pitch accent to signal the statement/question contrast. For example, Dimperio and House (1997) showed that in Neapolian Italian both questions and contrastive statements are characterized by a similar rise fall contour. However, the two contours differ in the timing of the high pitch target which is later in questions than statements. In addition to alignment, many languages distinguish questions from statements by means of pitch scaling. For example, in Bari Italian (Savino and Grice, 2007) and Hungarian (Gosy and Terken, 1994) differences of local scaling effects in the nuclear pitch accent have been shown to be relevant to perceptually cue sentence modality.
The interrogative syntax of polar questions in Persian is marked morphologically with a question particle 'aja'. However, questions are often signaled intonationally in Persian spontaneous speech. The intonational differentiation of questions and statements has been attributed to variation in the utterance-final F0 movement, namely IP boundary tone: H% in questions and L% in statements. This study was intended to investigate the possible phonetic exponents of the statement/question contrast in Persian, in particular, the variables that have turned out to be relevant in previous research.

2- The current study

The present paper aims at comparing the intonation of Persian statements and questions in terms of the pitch scaling and alignment of tonal events in pre-nuclear and nuclear pitch accents as well as utterance final pitch movement. The objectives of this research involve two production experiments. Experiment 1 compares the alignment and scaling patterns of pre-nuclear and nuclear rising accents in statements and questions. It is hypothesized that the kind of phonetic adjustments speakers make when producing comparing sentences under different sentence mode conditions can systematically affect the realization of both the nuclear and pre-nuclear accentual gestures. Experiment 2 examines the phonetic realization of the low phrase accent and the following H% boundary tone in questions. In this experiment, I particularly examine the effects of tonal crowding and question length on the scaling and alignment of the low plateau between the nuclear accent peak and the final rise. It is hypothesized that the surface scaling and alignment patterns of tonal targets are sensitive to the tonal crowding conditions in which they are involved.
A corpus of real speech with short declarative sentences was designed in order to elicit statement and question contours. The experimental sentences contained the word order (SUB)-OBJ-VERB-ADV (PP). The target words were the objects and verbs, both embedded in non-final position to avoid intonational pre-boundary effects. 16 speakers of standard Persian participated in the experiments. The speakers were all from Tehran, aged between 21 and 43, and were university teachers or students.

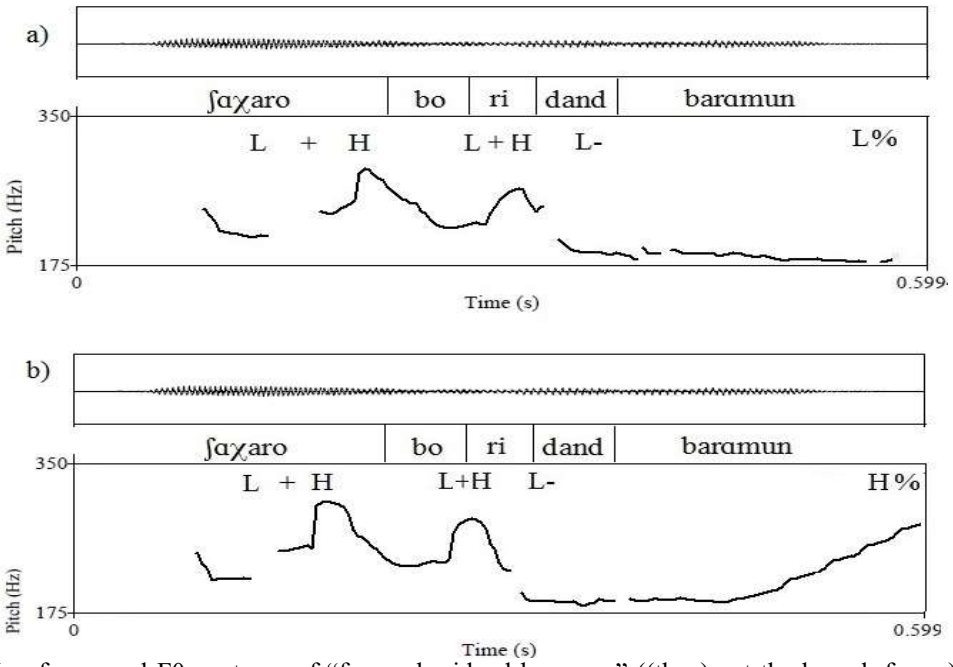


Figure 1: Waveforms and F0 contours of "ʃɑχaro boridand barɑmun" ((they) cut the branch for us), read as a statement (a) and a question (b) by Speaker MH.

For each target sentence, the experimenter initially recorded a context eliciting statement and question, spoken by a formal speaker. The contexts were presented together with the related target sentence, both visually on screen and auditorily over headphones. Speakers were asked to listen to the context and then read the answer as a response to the context question.

3- Results

Results of Experiment 1 revealed that questions have a higher pitch register than statements, manifested as an overall higher pitch level and a wider pitch span. Results of this experiment further revealed that the accentual peak of both pre-nuclear and nuclear pitch accents is realized higher and later in questions than statements. Results of Experiment 2 showed that the exact contour shape of questions in Persian depends on the length of the question, and that the position of the first peak and the low plateau depends on the position of the stressed syllables, and shows predictable adjustments in alignment depending on the proximity of adjacent tonal targets. The findings of this research follow easily from an autosegmental-metrical approach to intonational phonology, according to which melodies may contain long F0 stretches derived by interpolation between specified targets associated with metrically strong syllables and prosodic boundaries.

References

D'Imperio, M. (2000). The role of perception in defining tonal targets and their alignment. Ph.D. thesis, The Ohio State University, Columbus, OH.

Petrone, C., & Niebuhr, O. (2014). On the intonation of German intonation questions: The role of the prenuclear region. *Language and Speech*, *57*(1), 108-146.

Pierrehumbert, J., & Hirschberg, J. B. (1990). The meaning of intonational contours in the interpretation of discourse. In *Intentions in communication* (pp. 271-311).

Savino, M., & Grice, M. (2007, August). The role of pitch range in realising pragmatic contrasts–The case of two question types in Italian. In *Proceedings of the XVIth International Congress of Phonetic Sciences* (pp. 1037-1040). Saarbrücken, Germany: Pirrot GmbH.

Structural parallelism in Turkic, Iranian, and Uralic languages from a contact-linguistic perspective

VERONIKA MILANOVA[1], JEREMY BRADLEY[1], SAMPSA HOLOPAINEN[2] & ANDRÁS CZENTNÁR[3]
[1]University of Vienna; [2]University of Helsinki; [3]Eötvös Loránd University Budapest

Both the Volga-Kama Region of European Russia and Central and Southern Asia have been described as hotspots of linguistic convergence. In Central Asia, contacts between Uzbek and neighbouring Iranian varieties (Tajik, Dari) have led to well-established mutual influence (**Doerfer 1967**). In the Volga-Kama Region, the linguistic convergence resulting from long-standing language contact between Turkic languages (Tatar, Bashkir, Chuvash) and Uralic languages (Mari, Mordvinic, Permic) is frequently subsumed in the *Volga-Kama Sprachbund* (**Helimski 2003**, **Wintschalek 1993**, **Agyagási 2019**). Further, all languages of the regions have historically been subject to direct and indirect Iranian and Arabic influence (**Fragner 1999**, **Holopainen 2019**). While specific contact situations in the regions have long been subject to research, a holistic view has yet to be produced, and contact influences in the domain of morphosyntax and lexical typology remain comparatively understudied. The paper at hand offers a preliminary survey on a number of parallelisms in these two contact situations, as well as a number of conspicuous missing ones.

Aspectual auxiliary constructions: These structures, typical of Turkic languages (**Bradley 2016: 14**), consist of a converb and a superordinate verb which can be freely conjugated. The converb carries the semantic core of the pairing, while the second verb loses its lexical meaning partially or completely and primarily contributes an aspectual value. These structures have been widely adopted in a number of Uralic languages: Mari (**ibid.**), Udmurt (**Horváth 2013**), and also in two Samoyedic languages of Siberia, Kamass (**Klumpp 2002**) and Selkup (**Harder 2018**). They have also been observed in Tajik, and attributed to Uzbek influence (**Perry 2005: 224**).

(1) Mari (**Bradley 2016: 38**)

mal-en	kolt-aš
sleep-CVB	let_go-INF
‘to fall asleep.’	

(2) Tajik (**Perry 2005: 224**)

rafta	istoda-am
go.PTCP.PST	stand.PTCP.PST.-1SG
‘I am going.’	

Light verb constructions: The usage of light verbs in combination with nominal stems is typical of Iranian languages at large (**Korn 2013**), also borrowed into the Hindustani language (**Kuczkiewicz-Fraś 2003: 74–75**). In these a nominal loses its grammatical properties (case, number, etc. – importantly, they are not objects of the light verb) and the verb its semantics (i.e. it becomes semantically “light”). This pattern is typical of Turkic languages as well, and can be widely found in Udmurt (where the verb *karʲiɲi* 'to do' is itself an Iranian loan, cf. **Holopainen 2019: 380–381**), but is not common in Mari – the otherwise most heavily Turkified Uralic languages. Given the general global frequency of light verb constructions, it is difficult to ascertain the exact manner in which these structures spread.

- Persian *āwaz xwāndan* ‘to sing’ (lit. ‘song read’)
- Tatar *xis it-* ’to feel, to perceive’ (lit. ‘perception do’)
- Udmurt *keńeš karʲiɲi* ‘to consult’ (lit. ‘advice do’)

Reduplication: A number of reduplication patterns have been borrowed between languages in the regions under consideration. For example, in the so-called p-reduplication, the first syllable of an adjective serves as a reduplicant, it is prefixed to the original term with the final consonant replaced with a /p/ to create an emphatic form, e.g. Uzbek *qora* ‘black’ > *qop-qora* ‘pitch black’. This pattern can be found in numerous languages that have been in contact with

Turkic languages, such as Tajik *siyah* 'black' > *sip-siyah* 'pitch black', Mongolian *ulan* 'red' > *ub ulan* 'flaming red', Arabkir Armenian *nor* 'new' > *nop'-nor* 'brand new' (**Southern 2005: 79–85**), Mari *šeme* 'black' > *šep-šeme*.

Coordinating compounds: Coordinating compounds consisting of two elements of a class can serve as a collective noun referring to the class as a whole can be observed in the languages of the regions, e.g. Komi *paśkęm* 'clothing' < 'fur_coat-footwear', Chuvash *jĭvəś-kurək* 'vegetation' < 'tree-grass', Persian *kot-o-šalwār* 'suit' < 'jacket and trousers'. Furthermore, one can observe compounds consisting of two synonyms or near-synonyms with semantics that are difficult to ascertain, e.g. Farsi *ādāb-o-rosum* 'traditions' < 'manners and tradition' or Mari *šürgyvylyš* 'face' < 'face-ear'.
An open question is the factors that determined the similarity and difference of language contact in these two contact situations: which role did it play that there is a good typological fit (cf. **Thomason & Kaufman 1988: 74–76**) between the agglutinative Turkic and Uralic languages, but not between agglutinative Turkic and synthetic Iranian? Which factor did differing sociolinguistic facets of the respective contact situations play? Our presentation can be understood as an attempt to better delimit linguistic facets that would provide data to future surveys in this domain.

References

Agyagási, Klára. 2019. *Chuvash Historical Phonetics*. Turcologica 117. Wiesbaden: Harrassowitz.
Bradley, Jeremy. 2016. *Mari converb constructions: productivity and regional variance*. [manuscript of PhD thesis] Vienna: University of Vienna.
Doerfer, Gerhard. 1967. *Türkische Lehnwörter im Tadschikischen*. Wiesbaden: Franz Steiner Verlag.
Fragner, Bert G. 1999. *Die „Persophonie": Regionalität, Identität und Sprachkontakt in der Geschichte Asiens*. Berlin: Das Arabische Buch.
Harder, Anja. 2018. Converbal auxiliary constructions in Selkup. Presentation at the *7th International Conference on Samoyed Studies*.
Helimski, Eugene. 2003. Areal groupings (Sprachbünde) within and across the borders of the Uralic language family: A survey. *Nyelvtudományi Közlemények* 100. 156–167.
Holopainen, Sampsa. 2019. *Indo-Iranian borrowings in Uralic: Critical overview of sound substitutions and distribution criterion*. Helsinki: University of Helsinki.
Horváth, Laura. 2013. On the aspectual markers of the Udmurt language – Expressions of aspects in dialects. In: Márta Csepregi, Kata Kubínyi, Jari Sivonen (eds). *Grammmatika és kontextus – új szempontok az uráli nyelvek kutatásában III*. Budapest: ELTE. 108–123.
Korn, Agnes. 2013. Looking for the middle way: Voice and transitivity in complex predicates in Iranian. *Lingua* 135. 30–55.
Klumpp, Gerson. 2002. *Konverbkonstruktionen im Kamassischen*. Wiesbaden: Harrassowitz.
Kuczkiewicz-Fraś, Agnieszka. 2003. *Perso-Arabic Hybrids in Hindi*. New Delhi: Manohar.
Perry, John R. 2005. *A Tajik Persian Reference Grammar*. Leiden: Brill.
Southern, Mark. 2005. *Contagious Couplings: Transmission of Expressives in Yiddish Echo Phrases*. Westport, CT: Praeger.
Thomason, Sarah Grey & Terrence Kaufman. 1988. *Language contact, creolization, and genetic linguistics*. Berkeley, CA: University of California Press.
Wintschalek, Walter. 1993. *Die Areallinguistik am Beispiel syntaktischer Übereinstimmungen im Wolga-Kama-Areal*. Wiesbaden: Harrassowitz.

The rise of contrastive-partitive and definite functions of *=eš* in colloquial Persian: A grammaticalization from third person clitic pronoun

ZAHRA ETEBARI
Ferdowsi University of Mashhad

This talk concerns the development of the third person singular clitic pronoun *=eš* into denoting contrastive-partitive and definite functions in colloquial Persian. Possessive construction in Persian is regarded as pertensive or head-marked, since it is the head of the construction, the possessed noun, that is marked. One way of expressing the possessor in Persian is to attach a clitic pronoun to the possessed item. However, there are adnominal possessives in colloquial Persian in which the possessive *=eš* cannot be regarded as possessor, meaning that it does not refer to any lexical noun present either in the syntactic structure or in the context. One example is (1) in which *=eš* does not have any obvious reference, instead, its function is to mark the host as a part and contrast it with other parts of a similar set. This function is called contrastive-partitive, following Kiss (2018) who previously used the term for Hungarian. Similarly, in example (2), *=eš* does not refer to any third person lexical item. Here, *=eš* marks the noun in order to make it identifiable for the hearer; Hence, its function is considered as definite. This study has two goals: first, to give a description of the two functions in various authentic colloquial examples attained from internet chats, and second, to provide the grammaticalization pathways of the functions' development from the referential possessive.

These developments of possessive affixes or clitics have been attested in a few languages belonging to the Uralic (see e.g. Nikolaev 2003), Turkic (see e.g. Schroeder 1999), Austronesian (see e.g. Buren 2010), and Semitic (see e.g. Rubin 2010) language families. It will be argued that the possessive *=eš* in Persian, portrays a similar grammaticalization path which is resulted from a similar syntactic structure of possessive, i.e. pertensive and cross-referencing, in these languages. Works such as Jahanpanah (2001) and Naghzguy-Kohan (2014) have mentioned some of these non-possessive uses of *=eš* without providing a thorough explanation or describing their paths of grammaticalization. Moreover, their interpretations of the functions, considering contrastive-partitive *=eš* as an adverbial or an emphasis marker are to be argued against in this study.

The current work shows that the same nouns or adjectives bearing *=eš* with contrastive-partitive function can be used not only as adverbials, but also as subject and object arguments. Jahanpanah (2001) claims that the presence of *=eš* in some clauses adds an "even" meaning to it. Examples discussed in this talk indicate that the additional meaning is not always present and in some examples one can instead infer "only" from the use of *=eš*. It will be shown that the expectedness conception plays a role in the presence of the additional meanings. Fraurud (1999) points out that definite markers developed from the possessives in above mentioned languages are mostly used for associative anaphora. Nevertheless, in Persian, *=eš* with definite function is seen not only in deictic and associative anaphoric uses, but there are also examples of direct (non-associative) anaphoric use, which indicates that it has been more grammaticalized.

Clitic pronouns occur mostly in colloquial Persian and we expect to find the non-possessive functions in this mode, especially for the contrastive-partitive function which is motivated by pragmatic inference of contrast. However, to shed light on the path of grammaticalization that *=eš* may have taken, a historical corpus of New Persian has been searched. The study shows that the contrastive-partitive function has an origin in cross-

referencing possessive constructions with whole-part relation. In its grammaticalization path, *=eš* has been used, first, as a partitive marker – which is present in both colloquial and formal Persian– in cross-referencing structures where the lexical possessor is also present, and then lost its referentiality and extended to convey contrast as well — a function seen mostly in colloquial Persian. Furthermore, I argue that contrast marking of only one part makes it more emphatic, leading to development of additional meanings, "even" and "only". Investigation of the historical data shows that the source for the rise of definite function is associative possessives which date back to 15th century. The grammaticalization path provided in this talk demonstrates that *=eš* first has been used to denote associative anaphora and then developed further in the grammaticalization path to be used for (non-associative) direct anaphora, as a definite marker.

(1) *az* *bisavād=ā=**š*** *be-gir* *tā* *tahsilkarde-hā=ye*
from illiterate=PL=**PC.3SG** IMP-take.PRS to educated-PL=EZ
*taxasosi=**š*** *hičkodum* *bardāšt=ešun* *mesl=e*
specialized=**PC.3SG** none conception=PC.3PL like=EZ
ham *nist-Ø*
together NEG.COP.PRS-3SG
"From the illiterate to the educated specialized, everyone has her/his own conception."
(https://facebook.com/pg/mehrdadarghavani/posts, posted 31 October 2014)

(2) *moarefi=ye* *ketāb=e* *qodrat=e* *modiriyat=e* *zamān:* *xeyli*
introduction=EZ book=EZ ability=EZ management=EZ time very
*ketāb=**eš*** *xub-e*
book=**PC.3SG** good-COP.PRS.3SG
"An introduction to the book 'The ability of time management': The book is very good."
(https://mranderson.persianblog.ir, posted 10 Jan 2015)

References

Buren, Aaron D. 2010. The development of the Amharic definite article and an Indonesian parallel. *Journal of Semitic Studies*. 55 (1). 103-114.

Ewing, Michael. 1995. Two Pathways to identifiability in Cirebon Javanese. *Special Session on Discourse in Southeast Asian Languages of Berkeley Linguistics Society*. 21. 72-82.

Fraurud, Kari. 2001. Possessives with extensive use: A source of definite articles? In Irene Baron, Michael Herslund & Finn Sorensen (eds.), *Dimensions of Possession*, 243-267. Amsterdam-Philadelphia: John Benjamins *Publishing* Company.

Jahanpanah, Simindokht. 2001. Bound pronoun *=eš* and *dāštan*, two new tendencies in colloquial Persian of Tehran (Zamir-e mottasel-e *=eš* va *dāštan*: do gerāyeš-e tāze dar zābān-e fārsi-e goftāri-e emruz-e Tehrān). *Zabanshenasi*. 31. 19-43.

Kiss, Katalin E. 2018. Possessive agreement into a derivational suffix. In Bartos, Huba, Marcel den Dikken, Zoltán Banreti & Tamás Varadi (eds.), *Boundaries Crossed, at the Interfaces of Morphosyntax, Phonology, Pragmatics and Semantics*, 87-105. Cham: Springer International Publishing.

Naghzguy-Kohan, Mehrdad. 2014. From a clitic to a derivational affix (az vājebast tā vand-e ešteqāqi). *Ninth Conference on Iran linguistics* (Collection of papers). 332. 1345-1350.

Nikolaeva, Irina. 2003. Possessive affixes in the pragmatic structuring of the utterance: evidence from Uralic. In Suihkonen, Pirkko M.& Bernard Comrie (eds.), *International symposium on deictic systems and quantification in languages spoken in Europe and North and Central Asia* (Collection of papers), 130–145. Izhevsk and Leipzig: Udmurt State University and Max Planck Institute for Evolutionary Anthropology.

Schroeder, Christoph. 1999. *The Turkish Nominal phrase in Spoken Discourse*. Wiesbaden: Harrassowitz Verlag.

Posters

Linguistic Corpora of Iranian Languages. The Tajik National Corpus

ARSENIY VYDRIN
Institute for Linguistic Studies, St. Petersburg, Russia

Linguistic corpus is a collection of linguistically annotated texts[9]. There are different levels of linguistic annotation, for instance, syntactic, morphological, semantic or grammatical annotations. The texts in a corpus usually represent different genres (for example, fiction, poetry, newspaper, dialogues etc.) and time reference (modern or old texts). Linguistic corpus may include written (published) or oral (transcribed audio or video records) texts. One of the most important features of a linguistic corpus is a developed search system which allows to search by lexeme, translation, grammatical, syntactic, semantic or other tags, to search several items with a distance between them, to search information in certain texts or genres. The size of written corpora starts from a few million words while oral corpora are typically 10-100 times smaller. Linguistic corpora are usually available online. I will call a linguistic corpus a corpus which meets all the above-mentioned requirements.

Advantages of a linguistic corpus are uncountable. It is a unique tool which allows a scholar to obtain reliable language data in a very short time. Besides statistics, corpus gives an opportunity to work with big data and raise new tasks. Recently, the creation of a corpus became one of the main tasks of the language documentation.

A number of linguistic corpora were created during the last 30 years. To name just a few of the most known ones: British National Corpus, Czech National Corpus, Russian National Corpus. The importance of building linguistic corpora of modern Iranian languages was understood by the linguists years ago. Cf., for example, Gautier 1998. Building a Kurdish language corpus. Paper presented at ICEMCO 98 6th International Conference and Exhibition on Multilingual Computing. Cambridge, April 1998 (http://ggautierk.free.fr/e/icem_98.htm). However, as for now the majority of the modern Iranian languages do not have linguistic corpora. There are several corpora of Persian (e.g. Bijankhan, Hamshahri, Uppsala Persian Corpus) and Tajik (TajikWaC, https://www.sketchengine.eu/corpora-and-languages/tajik-text-corpora/). However, none of them meet the above-mentioned requirements of a linguistic corpus. They do not have any advanced search tools and the annotation is basically limited to parts of speech.

There are collections of texts of the extinct Iranian languages. The most famous one is Titus (http://titus.uni-frankfurt.de). Texts have grammatical tags and sometimes translation. However, the search possibilities are quite limited, for example, there is no option to search by grammatical tags, morphological markers or to search several items with a distance between them.

There are also some oral corpora of modern Iranian languages. E.g. Ossetic oral corpus (https://ossetic-studies.org/en/texts) or Wakhi corpus (recently published by J. Obrtelova. Narrative Structure of Wakhi Oral Stories. Uppsala: Acta Universitatis Upsaliensis, 2017). However, according to the available data, the size of these oral corpora is small, and the search is also not available.

The only Iranian languages with a linguistic corpus (in above-mentioned understanding of a linguistic corpus) are Ossetic and Tajik. Ossetic has Ossetic National Corpus (12 million words, http://corpus.ossetic-studies.org) and the written corpus of Digor dialect of Ossetic language (2.3 million words, http://corpus-digor.ossetic-studies.org). Both corpora were created in 2012-2014. The Tajik National Corpus was developed in 2019. Below I will consider the peculiarities of the Tajik National Corpus.

[9] This study was funded by the grant № 19-012-00637 of Russian Foundation for Basic Research (RFBR), project leader - Arseniy Vydrin.

At the present the total size of the corpus is 12 million words. The corpus consists of modern texts in the Tajik language, published in the 20th and 21st centuries. The corpus comprises the following genres: fiction, poetry, drama, journalism, scientific and educational texts, (auto)biographies, religious literature, political and law texts. The corpus is available online, the access is free of charge (https://tajik-corpus.org).

The corpus is based on the automatic annotation by a morphological analyzer. The automatic analysis includes lemmatisation and morphological tagging. Lemmatisation process implies attributing dictionary form to each word form. In the Tajik National Corpus, each word form is also translated into Russian (according to Tajik-Russian Dictionary by M.B. Rahimi, L.V. Uspenskaya. Moscow, 1954) and English. Lemmatisation implies manual processing of the dictionary. Morphological tagging includes grammatical, syntactic and semantic information (tags) such as parts of speech, tense, mood, person, number, causative verbs, proper names, male or female proper names, number (singular, plural, Arabic plural), postposition *-po*, izafat, color, body parts etc. The rules of morphological tagging were also processed manually.

The Tajik National Corpus was annotated automatically by the morphological analyzer. This analyzer works using a manually created Tajik-Russian dictionary and a formalized description of the Tajik morphology. It annotates each Tajik word in the text by its lemma, lemma translation to Russian or English and grammatical tags. For example, the Tajik word *омаданд* is automatically annotated as *омад-анд* V.STEM-PST.3PL 'to come, to arrive'.

The automatic annotation has both advantages and disadvantages. It minimizes manual labour and allows to enlarge the corpus with new texts constantly (in 2020, the size of the corpus was doubled by adding 6 million words of new Tajik texts). However, the automatic annotation sometimes offers ambiguous analyses. For example, the same Tajik word *омаданд* has the second annotation as *омад-анд* N.STEM-SG.COP.3PL 'arrival, coming' which is wrong (*'they are arrival/coming'). Out-of-dictionary words are not annotated. Disadvantages of the automatic annotation can be improved manually. At present, 91% of all words are annotated, the rest 9% should be manually added to the dictionary.

The Tajik National Corpus has a system of search developed for linguists. One can search by word, lemma, grammatical feature, gloss or translation. For an advanced search one can combine several parameters for a search query. One can set a position of an item in the sentence or to search several elements with certain distance between them. The search gives an option to select a subcorpus, i.e. to search only in certain texts. For instance, one can search only in fiction texts published in 2005 or to search only in poetry, to search in all works of a certain author etc.

A developed system of search and provided translations of each word allows the use of the Tajik National Corpus by scholars who are not familiar with the Tajik language.

In conclusion, one may state that at the moment very few Iranian languages have a linguistic corpus (Ossetic and Tajik). It appears that the main reason for the failure to develop a linguistic corpus of written texts for the major modern Iranian languages is the Arabic alphabet which does not show a big part of morphology and creates a lot of homonyms.

The accentual system of Parachi and Ormuri and its role for placing them among Iranian languages

ARTEM TROFIMOV
School of Advanced Studies in the Humanities, RANEPA, Moscow, Russia

V. A. Dybo established that the system of Pashto accentuation reflects the system found in Indo-Iranian and late PIE, i.e. with barytone and oxytone nouns and verbs corresponding to Indo-Iranian barytones and oxytones respectevely (Dybo 1974; Dybo 1989).

Two Iranian languages resembling Pashto in this respect are Ormuri and Parachi. These languages have free dynamic stress, mobile in the former language and nonmobile in the latter.
Efimov investigated the accentuation of Ormuri in a separate paper (Efimov 1985). His results demonstrate that the system of stress in Ormuri strongly resembles Pashto and the language inherited the Indo-Iranian system in the main points. In addition, the researcher demonstrated available data in his monograph (Efimov 1986).

I add Parachi data and list some Pashto-Ormuri-Parachi noun correspondences from different morphological classes:

Barytona
Pashto áspa f. ‘horse’ ~ Orm. Kan. yāsp Log. yåsp, f. yā́spa ‘horse’ ~ Parachi ōsp ‘horse’ ~ Vedic áśvā- ‘horse’.
Aфг. wā́wra ‘snow’ ~ Orm. Kan. γoř Log. γoš ‘snow’ ~ Parachi γarp ‘snow’ ~ Vedic vápra ‘a rampart, earth-work, mud-wall, mound’.
Pashto gū́ta ‘finger’ ~ Orm. Kan. ngušt, Log. angóxt ‘finger’ ~ Parachi γošt ‘finger’ ~ OInd. aṅgúṣṭha- ‘thumb’.
Pashto wə́č f. wə́ča ‘dry’ ~ Orm. Kan. wyok Log. wuk ‘dry’ ~ Parachi hóškō ‘dry’ ~ Ved. śúṣka- ‘dry’.
Pashto mə́x ‘face’ ~ Orm. Kan. mux Log. mox ‘face’ ~ Parachi mox ‘face’ ~ Ved. múkha- ‘mouth, face’.

I. Oxytona
Pashto plā́r ‘отец’ ~ Orm. Kan. pye ~ piyé Log. pe ‘отец’ ~ Vedic pitár- ‘отец’.
Pashto špá ‘night’ ~ Orm. Kan. šyo / šyu Log. xŏ ‘night’ ~ Parachi xawā́n ‘night’ ~ Ved. kṣapā́- ‘night’.
Pashto trə́, tərə́ ‘paternal uncle’ ~ Orm. tā ‘paternal uncle’ ~ Parachi peté ‘paternal uncle’ ~ OInd. pitr̥vyá- ‘paternal uncle’.
Pashto calŏ́r ‘four’ ~ Orm. Kan. tsār Log. tsår ‘four’ ~ Parachi čōr / čohŏ́r ‘four’ ~ Vedic catvā́ras ‘four’;
Pashto nwí, niwí, nəwí ‘90’ ~ Ormuri Kan. Log. nawí ‘90’ ~ OInd. navatí ‘90’.

There are many other examples including verbs, nouns, adjectives and numerals (it is interesting that numerals represent the most significant group due to the polysyllabic nature of decades; such forms reliably exclude coincidences).

I think that the resemblance of accentual systems in all three languages has importance for the subgrouping. There are two opinions on the place of Ormuri and Parachi among the Iranian languages. G. Morgenstierne listed numerous examples of Parachi and Ormuri words having specific Eastern Iranian correspondences and thought they Parachi and Ormuri

belonged to the Southeastern Iranian group (Morgenstierne 1926: 27–36). According to V. A. Efimov (Efimov 1986: 9; Efimov 2011: 5–6), the following reasons should lead us to consider that Parachi and Ormuri are Western, rather than Eastern, Iranian languages:
1) Ir. *b- > Orm. b-; 2) Ir. *d- > Orm. d-; 3) Ir. *g- > Orm. g-; 4) Ir. *-č > Orm. ž/z (Log. ž, Kan. z).

Furthermore, Efimov adduces certain reflexes that fit with the positioning of Ormuri in the Northwestern subgroup of Western Iranian: "the reflexes of Old Iranian *z/*d, *s/*θ and *sp corresponding with Ormuri z, s, sp, respectively" and "the reflex of Ir. *θr/*θr as Log. š, Kan. ř" (Efimov 2011: 6–7).

The preservation of the traces of the original Indo-Iranian stress in Pashto, Parachi and Ormuri represents an archaism, not the shared innovation and cannot be itself a decisive argument proving that these languages are closely related to each other. However, there are two considerations making it very unlikely that Ormuri and Parachi are northwestern Iranian languages. First, all other Iranian languages almost completely lost traces of the Indo-Iranian stress (and especially Western ones). Second, Ormuri and Parachi should at least be in a long-term contact with Pashto since antiquity to preserve the traces of accentual system unlike such languages as Kurdish. Taking into consideration some lexical innovations shared by Pashto, Ormuri and Parachi it is reasonable to think that the two latter languages are more closely related to Pashto.

References

Dybo, V. A. 1974. Afganskoe udarenie i ego znachenie dl'a indoevropejskoj i balto-slav'anskoj akcentologii. I. Imennaja akcentuacija. In: T. M. Sudnik (ed.). Balto-slav'anskie issledovanija: 67–105. Moskva: Nauka.

Dybo, V. A. 1989. Afganskoe udarenie i ego znachenie dl'a indoevropejskoj i balto-slav'anskoj akcentologii. II. Glagol'naja akcentuacija. In: N. I. Tolstoj (ed.). Slav'anskij i balkanskij fol'klor. Prosodija: Sbornik statej: 106–147. Moskva: Nauka.

Efimov, V. A. 1985. Udarenie v yazyke Ormuri. In: Indoiranskoe yazykoznanie. Ezhegodnik 1981: 31-64. Moskva: Nauka.

Efimov, V. A. 1986. Jazyk ormuri v sinkhronnom i istoricheskom osveschenii. Moskva: Nauka.

Efimov, V. A. 2011. The Ormuri language in Past and Present. English translation edited by Joan L. G. Baart. FLI language and culture series (Volume 6). Forum for Language Initiatives.

Morgenstierne, Georg. 1926. Report on a linguistic mission to Afghanistan. Oslo: H Aschehoug & Co.

Optative or Subjunctive? Two mood systems of Azerbaijani Talysh

DINARA STEPINA
Institute for Linguistic Studies RAS/ European University at Saint Petersburg, Russia

As for many languages, mood is one of the most important tools, alongside modal particles, for expressing different modal meanings in all varieties of Talysh. Within the Tatic group of North West Iranian languages, Northern Talysh is reported to be one of the few languages with Subjunctive and Optative moods remaining distinct in all its dialects, except for Leriki, while Optative is lacking in some Southern Talysh varieties, Vafsi, Khoiny, etc. (Stilo 2019, 723-724). Thus, Masulei (a Southern Talysh dialect) only has the Subjunctive mood (Stilo, 2019, 723)

It should be mentioned here that traditionally four main dialects are distinguished in Northern Talysh of Azerbaijan, named after the largest regional cities – Lenkorani, Astarai, Leriki, and Masally[10]. According to Miller, Northern Talysh of Azerbaijan does actually have the two distinct moods (Miller 1953, 146-151), although it is not specified whether all the dialects have both of them. My findings suggest that at least dialects of Masally and Lerik clearly distinguish between Optative and Subjunctive, even though some varieties of the latter are reported to allow using Optative instead of Subjunctive (see Stilo 2019, 724 for examples from Leriki).

Tables 1, 2. Optative and subjunctive of *votey* "to say" (present stem *vot-*) in Masally and Leriki dialects

Optative

	Singular	**Plural**
1	*bı-vot-om*	*bı-vot-omon*
2	*bı-vot-oş*	*bı-vot-oşon*
3	*bı-vot-o*	*bı-vot-on*

Subjunctive

	Singular	**Plural**
1	*bı-vot-ım*	*bı-vot-əmon*
2	*bı-vot-ış*	*bı-vot-iyon*
3	*bı-vot-ı*	*bı-vot-ın*

Tables 3, 4. Subjunctive of *hıtey* "to sleep" (present stem *hıt-*) and *doy* 'to give' (present stem *də-*) in Lenkorani and Astarai dialects

	Singular	**Plural**
1	*bı-hıt-om*	*bı-hıt-əmon*
2	*bı-hıt-oş*	*bı-hıt-əyon*
3	*bı-hıt-o*	*bı-hıt-on*

	Singular	**Plural**
1	*bı-də-m*	*bı-də-mon*
2	*bı-də-ş*	*bı-də-yon*
3	*bı-də-ø*	*bı-dən*

What remains unmentioned by previous researchers, is the fact that two other large dialects of Northern Talysh in Azerbaijan – those of Lenkoran[11] and Astara, have the only type of non-indicative mood (the *new* Subjunctive). By *new* I mean that the Lenkorani and

[10] There are also several Northern Talysh dialects in the neighboring areas in Iran.
[11] Lenkorani was the variety in which attempts of the Talysh language codification were done in 1920's and 1930's.

Astarai Subjunctive is syntactically and semantically different from the Subjunctive of Leriki and Masally. This new Subjunctive covers almost all the uses of Subjunctive and Optative in the other two dialects, including pure Optative uses (e.g., curses and blessings).

While some dialects of Azerbaijani Talysh do actually have the two distinct moods, there seems to be some significant difference among them in this aspect. Considering Stilo's observations on the occasional use of Optative instead of Subjunctive in Leriki, I suppose that semantic shifts in the use of moods in Lenkorani and Astarai triggered merging of the two formerly distinct moods into the single mood. The talk will also include the brief semantic analysis of the two mood systems in the main Northern Talysh dialects.

Bibliography

Miller B. Talyšskij Jazyk. Moscow: Akademiya Nauk, 1953, 266 p.
Stilo D. 'The Caspian region and south Azerbaijan: Caspian and Tatic', *The Languages and Linguistics of Western Asia: An Areal Perspective*, 2019, v.6, pp. 659-824.

Adverbial Relations in Modern Persian language: A Typological Study

FARYAR AKHLAGHI
Research Institute for Cultural Heritage and Tourism, Tehran, Iran

In the present study, a functional-typological approach has been applied, in order to study adverbial relations in Modern Persian. The adverbial relation is one of the three types of subordination, the other two are complement relation and relative relation. In the functional-typological approach, subordination does not consider different formal features in languages and is merely known as a tool to build a cognitive relationship between the two 'events' or 'situations', one of which is 'dependent event or situation' and the other is 'the main event or situation'. The dependent situation does not have an independent nature and its existence is possible alongside the main event (Cristofaro, 2005: 1-3). In this regard, adverbial relations link two events together, as such, one of them (the dependent event) indicates the circumstances under which the main event takes place. The adverbial relations can be classified based on their meanings and functions into six types: purpose relations- indicating the purpose of bringing about the main event; temporal posteriority ("before" relations), temporal anteriority ("after" relations) and temporal overlap (when relations)-the dependent event serves as a temporal reference point for the main event; reality condition relations- indicating under which condition the main event can take place; and reason relations- expressing the reason why or in which way the main event happens (Thompson and Longacre 1985, Givon 1990, Kortmann 1997, Cristofaro 2005).

In typological study of adverbial relations, the dependent clause and the changes which are made in it as the result of the subordination process can be considered as a cross linguistic assessable criterion. In this context, it is intriguing to study in a given language whether the structure of the dependent clause is different from the structure of the independent predicative clause or not. The assessment of this issue is possible by two variables; one, the features of the verb form of the dependent clause and the other, the coding of participants .

The present research relies on the study of the features of the verb form of the dependent clause in all six mentioned adverbial relations. Verbs coding dependent events may differ in structure for example they might not display all of the categorical distinctions (tense, aspect, mood, and person) which are allowed to verbs coding independent events. If so, they are called 'deranked' and if they have the same structure as the independent clause verb they are called 'balanced' (Stassen, 1985: 79-83). Studies show that in Persian the balanced verb form has an indicative mood and the deranked verb form has a subjunctive mood (Akhlaghi 2010).

Based on the typological information regarding the usage of deranked vs. balanced verb forms in the adverbial relations of the world languages, the "Adverbial Deranking Hierarchy" has been proposed as follows:

Purpose > Before, After, When > Reality condition, Reason

This hierarchy implies that if for coding the dependent clause the deranked form is used on the hierarchy at any point, then the deranked form is used at all points to the left (Cristofaro 2005: 168)

The present study has been conducted based on natural data from Modern Persian language. To collect data for this research, firstly, based on the semantic categorization provided for adverbial relations, a list of the most common adverbial subordinators which in Modern Persian proceed the clauses to express the concepts of the adverbial relations were prepared according to the author's language instinct and with reference to a synonym encyclopedia. Then, each of the subordinators in this list was searched separately in the digital 'Persian Language Database' and their applications were recorded. To collect more data, the given subordinators were searched manually in some fictional proses and newspapers or on the internet, and the resulting sentences were added to the corpus.

The findings indicate that in Modern Persian both balanced and deranked verb forms are used in coding the adverbial clauses and this is revealed by indicative/subjunctive distinction respectively. In this language, the only strategy which is used in "purpose relations" with the typical subordinator /be mænzur-e ʔin ke/ is deranking. Modern Persian uses deranking strategy for coding "before relation" with subordinators /Gæbl ʔæz ʔin ke / and the like, but balancing strategy for "after relations" with typical subordinator /bæʔd ʔæz ʔin ke/. In "when relations" with the typical subordinator /væGti ke/ the usual strategy is balancing, although some cases of deranked verb forms were observed. All the instances of the reality condition with the subordinator /ʔægær/ in the corpus of the present study used deranked verb form and the verb form for the reason relations dependent clause with typical subordinator /ʧon/ were balanced. The results are summarized in the following table:

Table 1: Balancing/Deranking strategy in adverbial relations in Modern Persian

Adverbial relations type	**Typical subordinator***	**Translation**	**Balancing/Deranking Strategy**
Purpose relations	/be mænzur-e ʔin ke/	in order to	Deranked
Temporal posteriority (before relations)	/Gæbl ʔæz ʔin ke /	before	Deranked
Temporal anteriority (after relations)	/bæʔd ʔæz ʔin ke/	after	Balanced
Temporal overlap (when relations)	/væGti ke/	when	Balanced/ Deranked
Reality condition (Condition relations)	/ʔægær/	if	Deranked
Reason relations	/ʧon	because	Balanced

***** There are more than one subordinator for each adverbial relation in Modern Persian. The most typical of them in each case has been represented in the table .1.

According to the above results, adverbial relations in Modern Persian do not obey the "Adverbial Deranking Hierarchy" because in this language "after relation" that precedes "reality condition" in the hierarchy, uses balanced verb form while the reality condition uses deranked verb form. Persian Language in this regard does not support the trace of typological findings.

Some Sources:

Akhlaghi, F. 2010. "Typological study of complementation in modern Persian language". [Barresi-ye radeshenaxti-ye mottamemafzayi dar Farsi-ye emrouz], Dastoor (Grammer). 6th issue. pp. 51-100.

Cristofaro, S. 2005. *Subordination.* Oxford: Oxford University Press.

Givon, T. 1980. "The binding hierarchy and the typology of complements", *Studies in Language 4*, 333-77.

Kortmann, B. 1997. *Adverbial Subordinators: A Typology and History of Adverbial Subordinators Based on European Languages*. Berlin and New York: Mouton de Gruyter.

Stassen, L. 1985. Comparison and Universal Grammar. Oxford: Basil Blackwell.

Thompson, S.A. and R.E. Longacre. 1985. "Adverbial Clauses", in T. Shopen (Ed.) Language Typology and Syntactic Description, Vol:2, Cambridge University Press, pp.171-234

Horizontal alignment and tense/aspect-restricted differential object marking in Iranian languages

JENS FLEISCHHAUER
Department of General Linguistics, Heinrich-Heine University Düsseldorf

The talk investigates two typologically rare features, which are only attested in Iranian and a rather small number of genetically closely related (Indo-Aryan and Dardic) languages. I argue that the two Iranian languages (Dersim) Zaza and Pashto as well as the Indo-Aryan language Kashmiri show a typologically very unusual restriction of object marking. The languages under discussion combine a functionally-motivated object marking asymmetry (differential object marking, e.g. Bossong 1985) with a restriction of object marking to particular tenses/aspects. An example from Dersim Zaza (Northwestern Iranian) is shown in (1). Dersim Zaza shows animacy-based differential object marking; only animate object arguments require oblique case marking (1a), inanimates do not (1b). As the examples in (1c) and (d) show, object marking is restricted to the present tense; in the past tense animate and inanimate object arguments treated a like. Object marking interacts with gender as well as number but this does not crucially affect the issue under discussion.

(1) Dersim Zaza (Selcan 1998: 278f.)
a. *Televe* *malım-u* *vinen-o.*
pupil teacher-OBL.PL see.PRES-3SG
'The pupil is seeing the teachers.'
c. *Televe* *kıtav* *cên-o.*
pupil book take.PRES-3SG
'The pupil is taking the book.'
b. *Malım-u* *televe* *di.*
teacher-OBL.PL pupil see.PST
'The teachers saw the pupil.'
d. *Televe-y* *kıtav* *di.*
pupil-OBL.SG book see.PST
'The pupil saw the book.'

Tense/aspect-restricted DOM has only occasionally be mentioned in the literature on case marking asymmetries: de Hoop & Malchukov (2008) mentioning Kashmiri and Arkadiev (2008) presents a brief discussion of Zaza.

Usually, languages showing differential object marking (DOM) do not restrict the case marking asymmetry to a particular tense or aspect. A particular example from the Iranian language family is the Western Iranian language Balochi (another clear-cut example is Persian, e.g. Lazard 1992 but also see the discussion of various Iranian languages in Bossong 1985). Balochi shows a quite complex pattern of case marking (for details see Farrell 1995 as well as Korn 2008, 2009); relevant for the current discussion is only that the language shows person-based differential object marking (2). A third person object argument is realized in the direct case form (2a), whereas a local person (first or second) receives oblique case marking (2b). Note that Balochi also shows differential subject marking, restricting oblique case marking to non-local subject arguments of transitive verbs in the past tense.

(2) Balochi (Farrell 1995: 214, 216)
a. *jinik-a* *bəcık* *yes.*
girl-OBL boy hit.PST
'The girl hit the boy.'
b. *mən* *tə-ra* *gitt.*
1SG 2SG-OBL catch.PST
'I caught you.'

Balochi does not have a tense/aspect-based restriction of DOM, which licenses the emergence of a horizontal alignment pattern under specific conditions (i.e., non-local subject of a transitive past tense verb in combination with a local object argument):

(3) *bəcik-ã* *ma-ra* *dist.*
boy-OBL 1PL-OBL see.PST
'The boy saw us.' (Farrell 1996: 216)

Horizontal alignment – also called 'double-oblique' – is typologically vary rare and only attested in a number of Iranian languages (e.g. Payne 1980). This alignment pattern does not allow discriminating between the two arguments of a transitive predicate. Horizontal alignment results from the combination of tense/aspect-based differential subject marking (DSM) and functionally motivated DOM in languages having a small (basically binary) case system (Arkadiev 2009).

In the talk, I first present evidence for the claim that horizontal alignment is restricted to languages combining DSM with DOM and having a binary case system. In a second step, I argue that the restriction of DOM to particular tenses/aspects is a strategy of avoiding the emergence of horizontal alignment due to functionally motivated case marking asymmetries. Horizontal alignment cannot arise in Dersim Zaza, Pashto or Kashmiri from the combination of DSM and DOM since object case marking is restricted to grammatical contexts in which the subject is not case marked.

In the talk, I investigate the known instances of tense/aspect-restricted DOM and present a functional motivation for its existence along the lines sketched above.

References

Arkadiev, Peter. 2008. *Differential Argument Marking in Two-term Case Systems and its Implications for the General Theory of Case Marking*. In Helen de Hoop & Peter de Swart (eds.), *Differential Subject Marking*, 151–171. Berlin: Springer.

Arkadiev, Peter. 2009. Poor (two-term) case systems – Limits of neutralization. In Andrej Malchukov & Andrew Spencer (eds.), *The Oxford Handbook of Case*, 686–699. Oxford: Oxford University Press.

Bossong, Georg. 1985. *Empirische Universalienforschung. Differentielle Objektmarkierung in den neuiranischen Sprachen*. Tübingen: Gunter Narr.

de Hoop, Helen & Andrej Malchukov. 2008. Case-Marking Strategies. *Linguistic Inquiry* 39(4). 565–587.

Farrell, Tim. 1995. *Fading Ergativity? A Study of Ergativity in Balochi*. In Theodora Bynon, David Bennett & George Hewitt (eds.), *Subject, Voice and Ergativity*, 212–235. London: School of Oriental and African Studies, University of London.

Korn, Agnes. 2008. Marking of Arguments in Balochi Ergative and Mixed Constructions. In Simin Karimi, Vida Samiian & Donald Stilo (eds.). *Aspects of Iranian Linguistics*, 249–276. Cambridge: Cambridge Scholars Press.

Korn, Agnes. 2009. The Ergative System in Balochi from a Typological Perspective. *Iranian Journal of Applied Language Studies* 1 (1): 43–79.

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

Payne, John. 1980. The decay of ergativity in Pamir languages. *Lingua* 51. 147–186.

Selcan, Zülfü. 1998. *Grammatik der Zaza-Sprache*. Berlin: Wissenschaft & Technik Verlag.

'kampyutere shaxsi' or 'rayanehe shaxsi': Deep Learning Methods for the Extraction of Persian Multi-word Expressions with English Loanwords and their Persian Equivalents

KATARZYNA MARSZAŁEK-KOWALEWSKA
Adam Mickiewicz University, Poznan, Poland

Today, almost 18 years after the seminal paper "Multiword Expressions - A Pain in the Neck for NLP" by Sag et al. (2002), multi-word expressions (MWEs) are still interesting and challenging aspect of Natural Language Processing (NLP), which is reflected in the number of papers addressing this phenomenon as well as number of people contributing to, and attending workshops and conferences such as SIGLEX-MWE or PARSEME. MWEs are lexical items that can be decomposed into multiple component words, but have properties that are idiomatic, i.e., idiosyncratic or unpredictable, with respect to their component words (Baldwin and Kim, 2010). They are very frequent in language and range over a number of different linguistic constructions, from idioms, e.g. *to pay an arm and a leg*, to fixed expressions, e.g. *rock and roll*, light verbs, e.g. *take a shower*, to noun compounds, e.g. *golf club*. Biber et al. (1999) claim that the number of MWEs in spoken English is 30% - 45% and 21% in academic prose. Jackendoff (1997) suggests that the number of MWEs in a speaker's lexicon is the same as simple words, yet if we take into consideration the domain specific lexicons this number seems to be an underestimation (Sag et al. 2002). Indeed, the research conducted by Ramisch (2009) suggests that the MWEs ratio can be between 50% and 80% in a corpus of scientific biomedical abstracts. Krieger and Finatto (2004) estimate that MWEs can constitite more than 70% of specialized lexicon.

MWE identification is the task of finding multiword expressions in a running text. For a very long time, this task was considered very challenging, not to say unrealistic. Many early work on MWE identification have focused on distinguishing between the idiomatic and literal usages of MEW, e.g. *kick the bucket* with the meaning of 'die'. The most exploited approaches to MWE identification focused on rule-based matching, supervised classification, sequence tagging, and parsing. Only recently, we have observed first attempts to MWE identification using deep learning. First such system was presented by Klyueva et al. (2017) as PARSEME shared task and it influenced the development of more advanced systems (e.g. Gharbieh et al. 2018). Only a year later, 9 out of 17 systems submitted to PARSEME used neural models.

In this paper we propose the first deep learning model for identification of Persian MWEs. We focus on a very specific group: Persian MWEs that contain English loanwords and their Persian equivalents proposed by the Academy of Language and Literature. There is a growing number of English loanwords in Persian, for which the Academy proposes Persian equivalents, e.g. [varzesh] for 'sport', [xodaks] for 'selfie', [fanavar] for 'technology'. One way of studying these substitutions could be by identifying, extracting and comparing the MWEs in which they occur, i.e. comparing loanword MWEs ([elme kampyuter] 'computer science', [bazihaye kampyuter] 'computer games', [kamputere kwantum] 'quantum computer') with its Persian equivalent MWE ([saxtafzarye rayaneh] 'computer hardware', [abre rayaneh] 'supercomputer', [bazihaye rayaneh]'computer games'). The research on MWEs in Persian has so far focused mainly on verbal multi-word units and, on light verb constructions (LVCs) in particular (Fazly et al. 2007, Salehi et al. 2012, Salehi et al. 2016). Thus, to our knowledge, this is the first attempt to analyse Persian MWEs with English loanwords and their Persian counterparts using deep learning methods.

For the purpose of this study, we employ deep learning architectures. Specifically, after Gharbieh et al. (2017) we consider a layered feed-forward, a recurrent neural network and a convolutional neural network. We will present and compare different features used for the study (BIO annotations, POS tagging, word-shape features) as well as various representations: from classical token-level *word2vec*, character-based *fastText* and finally deep contextualized word representations like *ELMo* and *BERT*. Since, the multilingual-BERT (Devlin et al. 2019) suffers from OOV (out-of-vocabulary) problem, we trained our own BERT model for Persian (*PerBERT*), which to our knowledge is the first such attempt.

References

- Baldwin, Timothy and Su Nam Kim (2010). "Multiword Expressions". In: Handbook of Natural Language Processing, Second Edition. Ed. by Nitin Indurkhya and Fred J. Damerau, pp. 267–292.
- Bieber, Douglas and Susan Conrad (1999). "Lexical bundles in conversation and academic prose". In: Out of Corpora: Studies in Honour of Stig Johansson. Ed. by Hilde Hasselgård and Signe Oksefjell. Amsterdam: Rodopi, pp. 181–190.
- Devlin, Jacob, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. (2019). "BERT: Pre-training of deep bidirectional transformers for language understanding. In: Proceedings of NAACL.
- Fazly, Afsaneh, Suzanne Stevenson, and Ryan North (2007). "Automatically learning semantic knowledge about multiword predicates". In: Language Resources and Evaluation 41, pp. 61–89.
- Jackendoff, Ray (1997). "Twistin the night away". In: Language 73 (3), pp. 534–559.
- Gharbieh,Waseem, Virendrakumar C. Bhavsar and Paul Cook. 2017. "Deep Learning Models For Multiword Expression Identification". In: Proceedings of the 6th Joint Conference on Lexical and Computational Semantics (*SEM 2017). Vancouver: Association for Computational Linguistics, pp. 54-64.
- Klyueva, Natalia, Antoine Doucet and Milan Straka. 2017. "Neural Networks for Multi-Word Expression Detection". In: Proceedings of the 13th Workshop on Multiword Expressions (MWE 2017). Valencia: Association for Computational Linguistics, pp. 60-65.
- Krieger, Maria da Graça and Maria José Bocorny Finatto (2004). Introdução à terminologia: teoria & prática. Sao Paulo: Contexto.
- Ramisch, Carlos (2009). "Multi-word terminology extraction for domain-specific documents". Master Thesis. Grenoble, France: École Nationale Supérieure d'Informatique et de Mathématiques Appliquées.
- Sag Ivan A., Timothy Baldwin, Francis Bond, Ann A. Copestake and Dan Flickinger (2002). "Multiword Expressions: A Pain in the Neck for NLP". In: Proceedings of the 3rd International Conference on Intelligent Text Processing and Computational Linguistics. CICLing-2002, pp. 1–15.
- Salehi, Bahar, Narjes Askarian, and Afsaneh Fazly (2012). "Automatic Identification of Persian Light Verb Constructions". In: Computational Linguistics and Intelli- gent Text Processing. CICLing 2012. Lecture Notes in Computer Science. Berlin, Heidelberg: Springer.
- Salehi, Bahar, Paul Cook, and Timothy Baldwin (2016). "Determining the Multi- word Expression Inventory of a Surprise Language". In: Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers. Osaka, Japan: The COLING 2016 Organizing Committee, pp. 471–481.

Definiteness markers in New Western Iranian languages, from a typology perspective

MARYAM NOURZAEI
Linguistics, Otto-Friedrich-Universität Bamberg, Germany

The aim of the present paper is to explore the development of grammatical markers for definiteness in New Western Iranian languages of the Zagros Mountains, which are unique among Iranian languages in this regard. The paper focuses on a sample of six Iranian languages: Balochi, Koroshi, Northern Kurdish, Lori, Spoken New Persian and Shirazi. The data for the present paper was extracted from six short narrative texts based on a questionnaire and published corpora for Balochi and Koroshi (Nourzaei et all 2015, Nourzaei 2017 and Nourzaei forthcoming) and Northern Kurdish (Öpengin 2016).

While the grammaticalization of definite markers has been a central issue in grammaticalization theory, they regularly cited cases (e.g. Romance) involve the development of erstwhile demonstratives, or linking particles such as relatives, into articles (Lyons 1999 and Himmelmann 2001). In Iranian languages under consideration, there is no obvious candidate for these markers among the known Iranian demonstratives or linking particles. More interesting, we witness a distinct development whereby an original derivational suffix, with a diminutive sense, changes into a suffixal marker of discourse identifiability. The question is that building this likelihood that these markers do not originate from demonstratives or linking particles in these languages but a diminutive: what does the functional distribution of these elements in these languages reveal regarding historical processes in the grammaticalization of definiteness?

The paper concludes that there is no obvious candidate for these markers among the known Iranian demonstratives or linking particles instead, the most plausible candidate appears to be an older diminutive suffix. This appears to be the first solidly attested case of development. In addition, the various markers (i.e. *-ak(a), -ū, -e, -ok and ak*) are all historically related, which allow us to reconstruct the following historical development for New Iranian definiteness: diminutive > endearment > proximity > identifiability > definiteness.

Examples:
Ex.1 Balochi
man ot-ī ***mard-ok-ā*** *dōst dār-īn*
'I love **my dear husband'** (UT)

Ex.2 Balochi
māmā man dōšī ***ges-ok-ā*** *māmel=a kort-on=o*
nok-en haptag=a raw-an be=te
'mother, I have bought **the house** and
we will move into it next week' (UT)

Ex.3 Koroshi
ar=ra's-ī be bāġ'bān-ā ar='raf-t
had=e ***ham=ī pīramar'd-ok-ā*** *a='š-ī*
he came (lit. arrived) to a gardener.
he went to **this old man**; he said ….' (Nourzaei 2017:568)

Ex.4 Shirazi
ye mard=ī o zan=ī mī-r-and bāzār mobl be-sūn-nan
A man and a woman go to the market to buy sofa.
***mard-ū** be **zan-ū** mī-g-e: sel kon…*
The man said to the woman: look,… (UT)

References:

Himmelmann, Nikolaus. 2001."Articles." In: M. Haspelmath, E. König, W. Oesterreicher and W. Raible (eds.). Language Typology and Language Universals. An International Handbook, Vol. 1. Berlin: de Gruyter, 831-841.

Lyons, Christopher. 1999. Definiteness. Cambridge university press

Nourzaei, Maryam, Carina Jahani, Erik Anonby & Abbas Ali Ahangar 2015. Koroshi. A Corpus-based Grammatical Description. Uppsala: Acta Universitatis Upsaliensis.

Nourzaei, Maryam. 2017. Participant Reference in Three Balochi Dialects: Male and Female Narrations of Folktales and Biographical Tales. Uppsala: Acta Universitatis Upsaliensis. (https://urn.kb.se/resolve?urn=urn:nbn:se:uu:diva-314090).

—Forthcoming. Documenting orality: Experiences of collecting folk narratives in Iran. Uppsala: Acta Universitatis Upsaliensis.

Öpengin, Ergin. 2016. The Mukri variety of Central Kurdish. Grammar, texts and lexicon. Wiesbaden: Harrassowitz.

Hybrid alignment patterns in ergative system of Iranian languages: evidence from Tākestāni

MEHRNOOSH TAHERKHANI[1] & FARNOOSH TAHERKHANI[2]
[1]University of Tehran, International Center for Persian Studies, Iran;
[2]Imam Khomeini International University - Qazvin, Iran

The Iranian languages are currently spoken across a vast stretch of Asia and their speakers inhabit several distinct geographic and cultural areas. Long-standing contact with numerous genetically diverse languages, among other factors, has led these languages to be typologically highly divergent; Yet despite the deep typological rifts cross-cutting the family, there is a striking grammatical property common to the vast majority of Iranian languages: the morphosyntax associated with past transitive verbs differs from that associated with all other verbs in the language concerned. This situation is often referred to as "split ergativity" (Dixon 1994) or as Haig (2008) puts it "Tense-Sensitive Alignment". The non-past tense clauses in these languages tend to follow a nominative-accusative case/agreement system (1a) and the past transitive clauses have retained the ergative-absolutive alignment system of the Middle Iranian (1b) (glosses adapted):

(1) a. ana Mæryæm-e mi-vin-ende.
3.PL.NOM.DIR Maryam.3.SG-OBL PRS-see-3.PL 'They see Maryam.'
b. æsb-an Liv-a bəxward-a
horse-PL.OBL leaf.F-NOM ate-3.SG.F 'The horses ate the leaf.' (Karimi, 1391:2)

However, this is not the case with alignment system of Tākestāni- a northwestern Iranian language- and this language has spawned a mixed litter of hybrid case systems; In past-stem transitive sentences, subject is in Direct case and direct object, contrary to many other members of the family, is in Oblique case (though not always morphologically presented by an overt marker) and there is a clitic attached to the object which cross-references to subject:

(2). ana Mæryæm-a=šon vind-Ø.
3.PL.NOM.DIR Maryam.3.SG-OBL=CLT.3.PL see.PST.STEM 'They saw Maryam.'

One argument in support of the claim for Oblique case of the object stems largely from kinship terms. Kinship terms in western Middle Iranian had distinct Oblique forms, containing an [r]. In Tākestāni this marker, representing "*ær*" or "*er*", has expanded its usage and it is now further used "with all nouns denoting a human being, except proper nouns", though not in all environments where an Oblique case would be expected. For instance, it is used with non-kinship terms only when they are modified by a genitive attribute (either by a personal pronoun or by a pronominal clitic):

(3) a. čeme ræfeq-**ær** xo
1.SG.GEN friend-OBL from 'from my friend'
b. ræfeq-**ær**=em xo
friend-OBL=CL.1.SG from 'from my friend'

Otherwise, the general Oblique marker *e* is used:

(4) ji mællem-**e** xo
DEM teacher-OBL from 'from this teacher.'

when there is a kinship term in the object position of a past transitive verb, the term always appears with this Oblique case marker illustrating that in this structure the object really does not bear Direct case:

(5) ešte=m xal-**ær** vin(d).
2.SG.GEN=CL.1.SG aunt-OBL meet.PST.STEM 'I met your aunt.'

This argument is further supported by the rich system of case distinctions on pronouns. The system of pronouns in Tākestāni has preserved the two-way opposition between Direct (subject) and Oblique (non-subject) case. When a pronoun appears in the direct object position, it is always in the Oblique form showing that the position is not actually the host for the Direct case:

(6) a. **ama** Pærviz-e=mon vin(d).
1.PL.NOM.DIR Parviz.3.SG.M-OBL=CLT.1.PL meet.PST.STEM 'we met Parviz.'
b. Ali **čoma**=š bebe.
Ali.3.SG.DIR 1.PL.OBL=CL.3.SG take.PST.STEM 'Ali took us.'

This shift in grammatical case of the object of past-stem transitive forms from Direct to Oblique (although remains to be confirmed) could be attributed to contact with dominant Persian language which follows nominative-accusative in all tenses. One very important consequence of this case shift in past-stem transitive constructions is to lower the functionality of pronominal clitics in distinguishing arguments. Rasekh mahand (2009) states that due to the lack of case marking on the arguments in Tati languages (including Tākestāni), pronominal clitics have developed a double-duty function: they distinguish subject from object, introducing their host as the object. If the discussion here is on the right track, then it corroborates that his generalization is not correct. Pronominal clitics in Tākestāni do not mark objects since objects are already Oblique-case-marked; They only identify their reference as the subject.

There is yet another argument favoring this claim. Pronominal clitics in Tākestāni co-occur with hosts of different categories and functions (nouns, nominal modifiers, adverbs of quantity, and the verb (7a-d) respectively) some of which clearly are not the object of the verb; A fact which clearly runs counter to the assumption that pronominal clitics are oblique-case-markers:

(7) a. ana ketab-e=**šon** bexen(d).
3.PL.NOM.DIR book.3.SG-OBL=CLT.3.PL read.PST.STEM 'they read (the) book.'
b. ana Ali=**šon** ketab bexen(d).
3.PL.NOM.DIR Ali.3.SG.M=CLT.3.PL book.3.SG-OBL read.PST.STEM
'they read Ali's book.'
c. qeqene=**šon** boxa(rd)-i.
a little=CLT.3.PL eat.PST.STEM-PERF 'they have not eaten much.'
d. mærdæk-e xo apærsæst=**em**.
man.3.SG.M-ABL from ask.PST.STEM=CLT.1.SG 'I asked the man.'

Moreover, there is a noteworthy exception to the generalization of Tense-Sensitive Alignment in this language where alignment with verb "*want*" patterns with past transitive verbs:

(8) a. æz ketab-e=**m** moGo-Ø.
1.SG.NOM.DIR book.3.SG-OBL=CL.1.SG want.PRES.STEM 'I want (the) book.'

Here the syntax associated with the verb is basically identical to that of the Past Transitive Construction (7a): the 'Wanter' is obligatorily cross-referenced via an enclitic pronoun (here=*m* and š), just as an A-past is also obligatorily co-referenced via a pronominal clitic. Likewise, in both constructions the verb shows no overt agreement. But crucially, unlike the Past Transitive Construction, (8) is in the present tense. Following Haig (2008) it has been shown that this resemblance is not accidental and could be traced in the concept of *Indirect Participation.* it has been discussed that some facets of the semantics of the verb have strong correlations with its syntactic behavior and with the interpretation of its arguments.

Selected references

Baker, M. C. (2012). On the relationship of object agreement and accusative case: Evidence from Amharic. *Linguistic Inquiry.* 43, 255-274.

Haig, Geoffrey. 2008. Alignment Change in Iranian Languages: A Construction Grammar Approach. Berlin & New York: Mouton de Gruyter.

Rasekh Mahand, M. (2009). "A typological perspective on the pronominal clitics in Tati Language". *Journal of Linguistic researches.* Vol. 1. Issue 1. Pp. 1-10 [In Persian].

Yarshater, E. (1969). *A Grammar of Southern Tati Dialect*, Mouton, The Hague.

A Dependency Grammar Approach to Persian Ezafe Construction

MINOO NASSAJIAN[1] & RAZIEH SHOJAEI[2]
[1]Sharif University of Technology, Iran; [2]Leipzig University, Germany

Persian Ezafe is regarded as one of the most controversial and challenging issues in different studies such as generative linguistics and Natural Language Processing (NLP) fields (see Karimi and Brame 1986; Samiian 1994; Ghomeshi 1983, 1997, Larson and Yamakido 2005; Kahnemuyipour 2000, 2014, 2016; among others for generative syntactic accounts and Megerdoomian 2000; Nojoumian 2011; Noferesti and Shamsfard 2014; Asgari et al. 2014 for computational accounts). Ezafe is as an unstressed morpheme (-e after consonants and –je after vowels) that links a head noun, head pronoun, head adjective, head preposition, or head adverb to their modifiers in a constituent that is called the Ezafe construction. It is recognized and pronounced but usually not written. So, this results in a high degree of ambiguity in Persian texts. The non-representation of this important marker also poses challenges for NLP tasks such as machine translation, POS tagging, Name-Entity Recognition, converting text to speech etc.

This paper explores the behavior of Ezafe marker and tries to extract simple rules to effectively recognize this marker in Persian text. As its theoretical framework, the present paper will focus on Dependency grammar (Tesnière 1953) and try to discuss Persian data in the light of this grammar. Dependency grammar can be adapted to suit languages with free or flexible word order (Covington: 1990 a,b, 1994) like Persian language. In addition, based on Covington (2001), both word-at-a-time operation and low memory usage properties of dependency parsing give dependency grammar a distinct advantage over Generative grammar in Natural Language Processing (NLP) tasks. This method of parsing plays a significant role in speeding up computer operations. The account we put forth presents interesting implications for certain principles of Dependency grammar.

Within this framework, first we take a close look at Ezafe distribution in Persian text. We use Uppsala Persian Dependency Corpus (Seraji: 2015) containg 6000 sentences annotated with 31 tags. The sentences are then analyzed by applying labeled directed graph, single-head constraint, and acyclicity constraints (proposed by Nivre 2005) to dependency graphs to derive Ezafe constructions syntactic representation. The constructions under study include nonverbal phrasal categories since verbs do not take Ezafe. Based on the extracted parsed trees, we provide a syntactic account for Persian Ezafe construction to correctly identify Ezafe markers.

Secondly, based on syntactic analyses, we formulate seven simple Ezafe insertion rules to cover all cases of its occurrence in the nonverbal phrasal categories. The Ezafe insertion rule 1 applies solely to the pre-modifiers of head nouns and pronouns (hame-***je*** cetɑbhɑ ‘all books’). The domain of application of Ezafe insertion rule 2 is zero nominal constituents (miz-***e*** ʔali ‘Ali’s table’). The Ezafe insertion rule 3 applies to head adjectives and properly inserts Ezafe after them (ʔɑbi-***je*** camranɟ ‘light blue’). By the application of this rule, Ezafe occurs between the head adjective and its post-modifier. Rules 4 and 5 respectively account for the occurrence of Ezafe between head prepositions and head adverbs and their post-modifiers (poʃt-***e*** miz ‘behind the table’ and hamɑnand-***e*** ʔin cetɑb ‘such as this book’). The domain of application of Ezafe insertion rule 6 is constituents that are not zero (lebɑs-***e*** sabz-***e*** camranɟ-***e*** ʔali ‘Ali’s light green shirt’). Finally, rule 7 is formulated to cover all cases of coordination (cetɑb-***e*** ɢermez-***e*** ʔali va cif-***e*** zard-***e*** Minoo ‘Ali’s red book and Minoo’s yellow bag’). This rule also covers Ezafe insertion in post-modifiers of noun heads.The

proposed rules can be used in different NLP tasks related to dependency parsing. The merit of our approach is that it circumvents the necessity of determining phrase borders in order to recognize Ezafe markers, opposing generative studies.

References

Covington, M. A. (1990a). "A dependency parser for variable-word-order languages". *Technical Report AI-1990-01*. University of Georgia.

Covington, M. A. (1990b). "Parsing discontinuous constituents in dependency grammar". *Computational Linguistics,* 16: 234–236.

Covington, M. A. (1994). "Discontinuous dependency parsing of free and fixed word order: Work in progress". *Technical Report AI-1994-02*. University of Georgia.

Covington, M. A. (2001). "A Fundamental Algorithm for Dependency Parsing". *Proceedings of the 39th Annual ACM Southeast Conference*: 1-8.

Ghomeshi, J. (1997). "Non-projecting Nouns and the Ezafe Construction in Persian". *Natural Language and Linguistic Theory,* 15: 729-788.

Kahnemuyipour, A. (2006). "Persian Ezafe construction revisited: Evidence for modifier phrase*". In Proceedings of the annual Conference of the Canadian Linguistic Association*: pp. 17–185.

Kahnemuyipour, A. (2014). "Revising the Persian Ezafe construction: A roll-up movement analysis*". Lingua*: pp. 1-24.

Kahnemuyipour, A. (2016). "The Ezafe Construction: Persian and Beyond*". ConCall-2, Indiana University*.

Karimi, S. and M., Brame. (1986). "A Generalization Concerning the Ezafe Construction in Persian". *Paper presented at Western Conference on Linguistics*: Canada.

Larson, R. and H., Yamakido. (2005). "Ezafe and the Deep Position of Nominal Modifiers*". Paper presented at the Barcelona Workshop on Adjectives and Adverbs*: Universitat Pompeu Fabra.

Nivre, J. (2005). Dependency grammar and Dependency parsing. Technical Report MSI report 05133, Växjö University: School of Mathematics and Systems Engineering.

Samiian, V. (1983). "Origins of Phrasal Categories in Persian, an X-bar Analysis". PhD Dissertation: UCLA.

Samiian, V. (1994). "The Ezafe Construction: Some Implications for the Theory of X-bar Syntax". In: Marashi, M. (Ed.). *Persian Studies in North America, Betheda, MD: Iranbooks*, 17-41.

Seraji, M. (2015). *Morphosyntactic Corpora and Tools for Persian*. Uppsala Universitet: Sweden.

Noferesti, S. and M., Shamsfard. (2014). "A Hybrid Algorithm for Recognizing the Position of Ezafe Constructions in Persian Texts". *International Journal of Interactive Multimedia and Artificial Intelligence* Vol. 2, N. 6:17-25.

Tesnière, L. (1953). *Esquisse d'une Syntaxe structural*. Paris: Klincksieck.

The corpus-based study post-predicate elements in spoken Persian

MOHAMMAD RASEKH-MAHAND & ELHAM IZADI
Linguistics Dapartment, Bu-Ali Sina University, Hamedan, Iran

In terms of word-order typology, Persian is regarded as an 'SOV' language. However, in spoken and informal written Persian, some of the constituents could be rearranged (Birner and Mahootian 1996, Karimi 2003, 2005, Rasekh-Mahand et al 2016). One representation of these rearrangements is the placement of certain constituents after the predicate. These constituents may have different syntactic roles:

1- *rafte* ***tabaqey*** ***bâlâ*** ***(weight:1, goal, inanim)***
go-3SG.PCPL floor=EZ up
'he had gone upstairs'

2- *Mehrâd bere* ***bâ*** ***dusteš (weight: 1, comitative, anim)***
Mehrâd SBJ-go.PRS-3SG with friend=3SG
'Mehrdad had gone with his friend'

3- *doyidam* ***tu*** ***kuče (weight: 1, location, inanim)***
run.PST-1SG in alley
'I ran out to the alley'

4- *hatman haminjur zade* ***beheš (weight: 1, poss-rec, anim)***
surely this like hit-PCPL to=3SG
'surely he had had a simple accident'

5- hiči umad **doktore (weight:0, dislocated topic, anim)**
EXCL come.PST-3SG doctor
'so the doctor came'

6- *parteš karde* ***čeqad unvartar mašin (weight: 2,*** **dislocated topic, anim)**
throw.PRS=3SG do.PRS-PCPL how much that side-COM car
'and thrown him several meters away'

Frommer (1981) is the first, study of post-verbal phenomenon in Persian. He has based his study in a small corpus, which includes formal and informal spoken and written varieties. However, his corpus is not good representative of colloquial spoken Persian. There are some other studies, too (Haig 2014, Haig and Theile 2014).. In this study, we try to explore and explain the placement of constituents after verb by using a corpus of natural spoken Persian (called HamBam; which is built by cooperation of Hamedan and Bamberg universities; some hours of real data of Persian speakers). We try to explain the different syntactic and pragmatic factors which force post-posing elements to post-verbal position in Persian (in framework of an excel file which shows the factors, frequency, positions and examples which is basically has prepared by the Bamberg University). Different factors such as grammatical role, syntactic heaviness, information structure, verb type, preposition type, definiteness and animacy are considered. The preliminary findings show that syntactic roles and verb type are decisive factors, while some of the other factors could force postposing, too. Some roles such as goal, addressee and recipient and location are in high frequency to postposed in spoken

Persian and in preliminary findings it has cleared that we have a lot of examples of these roles. The other factor is verb type, certainly most of these roles accompany with emotion verbs such as go, give, tell and arrive and also these verbs have high frequency in our corpus.

References

Arnold, J. E., Wasow, T., Losongco, A., & Ginstrom, R. (2000). Heaviness vs. newness: The effects of structural complexity and discourse status on constituent ordering. *Language*, 76, 28_55.

Birner, B., & Mahootian, S. (1996). Functional constraints on inversion in English and Farsi. *Language Sciences*, *18*(1-2), 127-138.

Comrie, B. (1989). *Language Universals and Linguistic Typology*, revised ed., Chicago: University of Chicago Press.

Croft, W. (2003).*Typology and universals*.2nd ed. Cambridge: Cambridge University Press.

Dabir-Moghaddam, M. (2001). "Word Order Typology of Iranian Languages". The Journal of Humanities 2.8, 17 -23.

Dryer, M.S. (1992). "The Greenbergian Word Order Correlations". *Language* 68, 81-138.

Frommer, P.B (1981). Post-verbal phenomena in colloquial Persian syntax. University of Southern California. Faculty of the Graduate School.

Greenberg, J. (1966). Language Universals: With Special Reference to Feature Hierarchies.The Hague: Mouton & Co.

Haig, G. (2014). Verb-Goal(VG) word order in Kurdish andNeo-Aramaic: typological and areal considerations. In: Khan, Geoffrey & Lidia Napiorkowska (eds.) *Neo-Aramaic and its linguistic context*. New York: Gorgias Press, 407-425

Haig, G. and Thiele, H. (2014). Post-predicate Goals in Northern Kurdish and neighbouring languages: a pilot study in quantitative areal linguistics. Presentation at the Second International Conference on Variation and Change in Kurdish, MardinArtuklu University, Turkey, 8-9th October 2014.

Karimi, S. (2005). A minimalist approach to scrambling: Evidence from Persian (Vol. 76).

Karimi, S.(2003). On object positions, specificity, and scrambling in Persian. In Karimi, S. (ed.), *Word order and scrambling.* Blackwell Publishing. 91–124.

Rasekh-Mahand, M., Alizadeh-Sahraie, M., & Izadifar, R. (2016). A corpus-based analysis of relative clause extraposition in Persian. *Ampersand*, *3*, 21-31

Definiteness marking in the languages of Chahar Mahal va Bakhtiari Province, Iran

MORTAZA TAHERI-ARDALI
Shahrekord University, Iran

Definiteness has been a central topic in modern linguistics since its advent in the late 19th century. In Christophersen's (1939) view, what distinguished definite from indefinite descriptions is whether or not the addressee of the utterance is presumed to be acquainted with the referent of the NP (Abbott 2004). This aspect of the topic is central, but the issue is more complex. A cross-linguistic investigation of this phenomenon is found in Haspelmath (1997), Lyons (1999) and Aguilar-Guevara et al. (2019). In Iranian linguistics, a few studies focusing on this topic have appeared, including Jahani (2015), Haig (2018) and Nourzaei (2019). Still, in most linguistic descriptions of Iran's languages, definiteness has been meagerly addressed. A cross-linguistic study of definiteness in linguistic areas within the country is lacking. This paper focuses on the ways in which languages in Chahar Mahal va Bakhtiari Province of Iran (hereafter C&B) express definiteness overtly.

Initial assessment shows that Bakhtiari, Charmahali, Turkic and Persian are the main vernaculars spoken in C&B Province (Taheri-Ardali & Anonby 2017). Bakhtiari as the dominant language is spoken in the north-west, west, south and central areas plus the provincial capital Shahr-e Kord. Charmahali, Turkic and Persian are spoken mainly in the north-east and eastern sections of the province. Moreover, Persian is extensively used as a lingua franca by almost all speakers (Taheri-Ardali et al. 2016).

In our study, linguistic data have been collected from 31 language varieties in 26 research sites over more than one year using a five-part questionnaire set developed for language variation in the *Atlas of the Languages of Iran* (ALI) project (Anonby et al. 2019). This paper tries to display how definiteness, as investigated through the morphosyntax section of the questionnaire, is encoded in the languages of C&B province.

Results show that definiteness can be explicitly expressed in Iranic languages of the region. The definite marker takes the forms *-e*, *-eke* and *-ke* in almost all Bakhtiari varieties (*mālam-e* 'the teacher'; *maalem-eke koǰe-ne* 'where is the teacher'; *maalem-ke ke ve mašad vay* 'the teacher that came from Mashhad') (Anonby & Asadi 2014). The form *-a*, *-eka* and *-ka* may be used when followed by the object-marking clitic *-ne* (*moalem-a-ne diy-om* 'I saw the teacher'; *hey siv-eka-ne xor-e* 'he is eating the apple'; *seyv-ka-n xa* 'he ate the apple'). The suffix *-aku* has been found in Bakhtiari of Chilteh Duderā in southernmost part of the province, an area which shares some features with Southern Lori of Boir Ahmadi (Taheri 2014). The two forms *-eka* and *-ika* have been attested in Bakhtiari of Boldāji and Juneqān which are bilingual cities with Turkic as the dominant language. Charmahali vernaculars (except Sheykh Shabān, which uses the form *-a*) consistently use *-e* as the definite marker similar to modern colloquial Persian (Paul 2019). In 5 out of 8 Turkic varieties, *ǰaġaz* is used as a definite marker (*moalem ǰaġaz* 'the teacher'). The occurrence of *ǰaġaz* as a definite marker can be attributable to the linguistic contact with Iranic languages in the region (Schreiber et al. 2017). In conclusion, a definiteness isogloss across the province can be drawn not only between Iranic and Turkic, but also between the Iranic varieties Bakhtiari and Charmahali: the suffix *-ke* and its allomorphs have been found in all Bakhtiari vernaculars investigated, but appear to be absent from Charmahali vernaculars.

Like in colloquial Persian (Jahani 2015), in all Iranic languages in C&B province the definite marker can be added to a noun phrase consisting of a noun plus an adjective. But in both

Persian and all Iranic varieties of C&B the ezafe which connects the adjective to the noun is deleted when the construction is marked with the definite marker (*moalem ǰevun-e* 'the young teacher').

A case of areal diffusion of morphology from Bakhtiari into Turkic is attested in Turkic of Naqneh as an unrelated language (*-ākki* in *moalemākki madrasa da* 'the teacher is at school'). This spread of the Iranic definite marking nominal suffix has also been reported in Turkic of Sonqor (*šär-äkä-sı-nı*) (poem-def-poss.3sg-acc) 'this certain/aforesaid poem of his', which is influenced by the neighboring Iranic languages Kurdish (Bulut 2019).

Typologically, in Bakhtiari the definite marker *-(e)ke* is attached to the right periphery of the noun after plural marker (*eney seyv-ā-ka=ne xor-e* 'he is eating the apples') which is in contrast to Kurdish *dost-ak-ān* (friend-def-pl) 'the friends' (Khan 2007, Haig & Khan 2019).

References:

Abbott, B. 2004. "Definiteness and indefiniteness". In Laurence R. Horn & Gregory Ward (eds.), *Handbook of pragmatics*, 122–149. Oxford: Blackwell.

Aguilar-Guevara, A., J. P, Loyo & V. V-R, Maldonado, eds. 2019. *Definiteness Across Languages*. Vol. 25. Berlin: Language Science Press.

Anonby, E. & A. Asadi. 2014. *Bakhtiari Studies. Phonology, Text, Lexicon* (Studia Iranica Upsaliensia 24). Uppsala: Acta Universitatis Upsaliensis.

Anonby, E., M. Taheri-Ardali & A. Hayes. 2019. "The Atlas of the Languages of Iran (ALI): A research overview". *Iranian studies* 52.1-2, 199-230. Online at: https://www.tandfonline.com/doi/full/10.1080/00210862.2019.1573135.

Bulut, C. 2019. "The Turkic varieties of Iran". In Geoffrey Haig & Geoffrey Khan, eds, *The Languages and Linguistics of Western Asia: An areal Perspective*, 398-444. Berlin: De Gruyter Mouton.

Christophersen, P. 1939. *The articles: A study of their theory and use in English*. Copenhagen: Munksgaard.

Haig, G. 2018. "Optional definiteness in Central Kurdish and Balochi. Conceptual and empirical issues". Presented at *Information Structure in Spoken Language Corpora 3: Discourse and Information Structure* workshop. In Münster, Germany, Dec 7-8, 2018.

Haig, G & G. Khan. 2019. "Introduction". In Geoffrey Haig & Geoffrey Khan, eds, *The Languages and Linguistics of Western Asia: An areal Perspective*, 1-29. Berlin: De Gruyter Mouton.

Haspelmath, M. 1997. *Indefinite pronouns*. Oxford: Oxford University Press.

Jahani, C. 2015. "On the Definite Marker in Modern Spoken Persian". Presented at *International Conference on Iranian Linguistics* (ICIL6), Ilia State University, Tbilisi, Georgia, June 23-26, 2015.

Khan, G. 2007. "Grammatical borrowing in North-Eastern Neo-Aramaic." In Yaron Matras & Jeanette Sakel (eds.), *Grammatical borrowing in cross-linguistic perspective*, 197–214. Berlin & Boston: De Gruyter Mouton.

Lyons, C. 1999. *Definiteness.* Cambridge: Cambridge University Press.

Nourzaei, M. 2019. "The emergence of definiteness in New Western Iranian languages: Extending the typology of definiteness". In proceedings of the *52nd Annual Meeting of the Societas Linguistica Europaea.* 21-24 August 2019, Leipzig University, Germany.

Paul, L. 2019. "Persian". In Geoffrey Haig & Geoffrey Khan, eds, *The Languages and Linguistics of Western Asia: An areal Perspective*, 569-624. Berlin: De Gruyter Mouton.

Schreiber, L., M. Taheri-Ardali., G. Haig & E. Anonby. 2017. "Contact-induced Change in Irano-Turkic Morphosyntax." Paper presented at *the Conference on Language Contact and Language Change in Western Asia*, Goethe University Frankfurt, 10-12 March, 2017.

Taheri, E. 2014. "*Sākhthā-ye ma'refesāz dar Lori Boirahmadi* [Definiteness in Boirahmadi]". *Iranian Journal of Language and Linguistics*. No 16, pp. 57-70.

Taheri-Ardali, M., et al. 2016. "Chahar Mahal va Bakhtiari Province". In: Anonby, Erik et al. (eds.), *Atlas of the Languages of Iran.* Ottawa: Geomatics and Cartographic Research Centre. http://iranatlas.net/index.html?module=module.language-distribution.chahar_mahal_va_bakhtiari#.

Taheri-Ardali, M. & E. Anonby. 2017. "Language Distribution in Chahar Mahal va Bakhtiari Province of Iran." 1st International Conference on Languages, Dialects and Linguistics (ICCILDL1), Pazhoheshgaran Andishmand Institute, Ahvaz, Iran, 2-3 February, 2017.

The derivation of light verbs in Aheli

MUHAMMED OURANG
University of New South Wales (UNSW), Sydney, Australia

In this paper we examine the derivation of the light verb construction (LVC) in Aheli, a dialect of Lari, an endangered Iranian language. The main aim of this work is to answer these questions: a) which verbs participate in light verb constructions, what are the word classes of non-verbal elements in LVCs that form complex predicates (CPr) and finally what makes the LVC constructions in Aheli different from those in other Iranian languages, especially Persian. As it is typically the case with LVC cross-linguistically (Butt, 2010), the Aheli LVC often consists of a non-verbal element (NV) and a light verb (LV). The non-verbal element may be a noun, an adjective, or a prepositional phrase whereas the light verbs may be one of the following: *(va:)kerdæ* 'to make', *va:vɒzæ* 'to become'/'to get', *zætæ* 'to hit', *ændæ* 'to come', *geretæ* 'to get', *va:xærdæ* 'to drink', *keʃezæ* 'to pull'(see (1)-(3)). The paper has three interconnected goals. First, using the semantic templates of Megerdoomian (2004), which was developed on the basis of data from Farsi, we show that the Aheli LVCs can be divided into three categories: (a) change of state verbs, (b) activity verbs, and (c) instrumental/locative verbs. Second, we discuss how the transitivity of a verb may be different when it functions as a main verb as opposed to when it is used in the LVC. For instance, *bʊlezæ* 'to cut' is a transitive verb when used as a main verb, but it becomes intransitive in the LVC: *kævi: bʊlezæ* 'to engage' (ex. (1)). In addition, some verbs like *geretæ* 'to get' or *keʃezæ* 'to pull' which are transitive verbs make intransitive light verbs in constructions like *tæʃ geretæ* 'to catch fire' (lit. '**fire to_catch**') (ex. (4)) or *dæs keʃezæ* 'to procrastinate' (lit. '**hand to_pull**'). Moreover, some verbs only appear in LVCs and cannot act as main verbs. *Va:vɒzæ* 'to become'/'to get', *va:kerdæ* 'to make' can be listed in this category. Third, we show that the meaning of a number of LVCs is not compositional, that is, it is not possible to predict the meaning of the LVC from its constituents: *keʃezæ* 'to pull' but *dæs keʃezæ* 'to procrastinate'. We conclude the paper by pointing out some important differences between the Aheli LVC and the relatively well-studied Farsi LVC (Karimi, 1997; Pantcheva, 2008; Megerdoomian, 2012, among others). Thus, for example, in Farsi *gu:l zædæn* 'to deceive' (lit. '**deception to_hit**') is the causative form of *gu:l xɒrdæn* 'to be deceived' (lit. '**deception to_eat**') while in Aheli *læpʊk zætæ* 'to deceive' (lit. '**deception to_hit**') is employed in passive sentence to denote the intransitive variant as in (5) and (6).

Examples

(1) *kævi:* 'engagement' + *bʊlezæ* 'to cut' = *kævi: bʊlezæ* 'to engage'
æ:mɑ: sævɑ: **kævi: æ-bʊl-em**
we tomorrow engagement IPFV-cut.NPST-1PL
'we engage tomorrow' (lit. we cut the engagement tomorrow).

(2) *ʃeri*: 'sweet' + *kerdæ* 'to make' = *ʃeri: kerdæ* 'to put make-up on sb'
mæʃɑ:tæ beɪn **ʃeri: ʃ=æ-kerd-e**
make_up_artist bride sweet 3SG=IPFV-do.PST-PTCP
'the make-up artist used to put make-up on bride'

(3) *æ* 'to' + *la:* 'loss' + *va:nesæ* 'to put' = *æ la: vanesæ* 'to waste'
mægæ pu:l **æ lɑ: vɑ:nes-eʃ**
always money to loss put.NPST-2SG
'you always waste the money' (lit. you always put the money on loss)

(4) *tæʃ* 'fire' + *geretæ* 'to catch' = *tæʃ geretæ* 'to catch fire'
mʊjɑ: **tæʃ ʊʃ=gere**
palm_tree fire 3SG=catch.PST
'the palm tree caught fire'

(5) mʊ m=æ ʊn **læpʊk zæt**
I 1SG-to he deception hit.PST
'I deceived him'

(6) ʊn ʃ=æ mʊ **læpʊk zæt**
I 3SG-to I deception hit.PST
'he deceived me'

References

Butt, M. (2010). The light verb jungle: Still hacking away. Complex predicates in cross-linguistic perspective. In M. Amberber, B. Baker, & M. Harvey, (Eds.), *Complex predicates: Cross-linguistic perspectives on event structure* (pp. 48-78). Cambridge: Cambridge University Press.

Karimi, S. (1997). Persian complex verbs: Idiomatic or compositional. *Lexicology*, 3, 273-318.

Megerdoomian, K. (2004, May). A semantic template for light verb constructions. In *Proceedings of the First Workshop on Persian Language and Computers*. Tehran University, Iran. May (pp. 25-26).

Megerdoomian, K. (2012). The status of the nominal in Persian complex predicates. *Natural Language & Linguistic Theory*, 30(1), 179-216.

Pantcheva, M. (2008). Noun preverbs in Persian complex predicates. *Papers on Language & Linguistics*, 19-45.

Code-switching Pattern in Persian: A Sociolinguistic Study in Urmia

NAFISEH TADAYYON CHAHARTAQ, JAVID FEREIDONI & FATEMEH MEHDIPOUR HALVAEE
[1]Department of Linguistics, Alzahra University, Vanak, Tehran, Iran;
[2]English Department, Urmia University of Medical Sciences, Urmia, Iran; [3]Free Persian Translator

1. Background and aims

Code-switching, or the alternation between two or more languages or linguistics varieties, is one of the possible outcomes of language contact and can be seen in most of the multilingual societies. One example of such communities is Urmia, a city in northwest of Iran with several linguistic varieties. Among these varieties are Turkic, the language of Turks who are the majority ethnic group; Farsi, the official language of the country, which also serves as a lingua franca among different ethnic groups; Kurdish, spoken by Kurds; Assyrian, used by Assyrians, and Armenian, spoken by Armenians. In this research, we aimed to comparatively study amount and direction (mother language to Farsi and vice versa) of code-switching among bilingual high-school students in Urmia.

2. Materials and methods

Our sample population included high-school male and female students from three different regions of Urmia each indicating low, middle, or high social class (6 high-schools, totally). The data included code-switched elements appearing in the students' daily talks. These elements were extracted from 6 hours (1 h per school) of conversation, collected through voice recording by a non-participant observation method. The code-switched utterances were selected based on a number of criteria which distinguished them from other language contact phenomena, e.g. borrowing. So an element was not considered as code-switching when: it was phonetically, morphologically, or syntactically integrated into the target language (e.g. Haugen 1973; Clyne 1987); there was an equivalence for it in the target language; it was commonly used in the speech community; it was a proper name, food name, or any other utterance socially integrated into the target language (e.g. Pfaf 1979, Poplack 1985). Here are some examples of code-switching from our data corpus (the code-switched elements are ***italic and bold***):

1- A. taqallob neveshti tu emtāhān ?

cheat write-2S on exam
"Did you cheat on the exam?"

B. Ki ? Man ? ***Kim yazıb buları manım qollarımda*** ?

who me *who wrote these my arms-on*
"Who? Me? Who wrote these on my arms?"

2- A. ... nemidunam , xeili ru asābame

NEG-know-1S. very on nerve-1S-is
"I don't know, he is really on my nerves."

B. boro , ***o bizim kılasımızın aşkısıdı , mage na*** ?

go *he our-GEN class-3Pl-GEN love-GEN-is if not*
"Don't say that, everyone in our class loves him, am I right?"

3- masxare***tarin*** sözdü ki eşıtmışam

ridiculous-*SUPER* word-is that hear-PAST-1S
"This is the most ridiculous thing I have ever heard".

4- Köküna beʃ sualdı nümrası ***punezdah***

totally five question-is mark-GEN *fifteen*
"Totally, there are five questions with 15 marks".

In first and second examples, one can see that speaker B alternates his code from Farsi to Turkish. In examples (3) and (4) the matrix language is Turkish, in which a Farsi element (superlative suffix –*tarin* in (2) and number *fifteen* in (3)) is embedded.

After extracting the data, they were analyzed using SPSS software version 16 through Independent samples t-test and ANOVA.

3. Findings

Our statistical findings showed that 90% of the code-switching happened when the dominant language of the conversation was the mother language. Besides, there was a statistically significant difference in behavior of high-class and middle- and low-class students regarding CS direction ($p<0.001$). To be more specific, students representing the high-class mostly switched from the official language to their mother language; while the reverse direction was dominant among low- and middle-class students. On the other hand, both sex groups acted similarly regarding both frequency and direction of code-switching ($p=0.92$ and $p = .13$, respectively); which means their code-switching pattern was similar, at least at the time of study.

4. Conclusion

To sum up, it can be inferred from the results that code-switching phenomena happens in the daily conversations of Urmia high school students as the outcome of language contact. Such a result was not far from expected, with the majority of the population being bi- or multilingual. Also, dominance of the official language in schools, its social prestige, and the linguistic situation in a multilingual city may reinforce this situation. Such a route, if continued, results in major changes in minority languages. However, more researches are required to be able to speak precisely.

References

Clyne, M. (1987). "Constraints on code switching: how universal are they?". *Linguistics*, 25, 739-764;

Fereidoni, J. (2003). "A sociolinguistic study on multilingualism: A domain analysis perspective". *Working Papers of the Linguistics Circle*, 17;

Gumperz, J. and Hernandez-Chavez, E. (1975). Cognitive aspects of bilingual communication. In Hernandez-Chavez, E., Cohen, A. and Beltramo, A. (Eds), *El Lenguajede fos Chicanos*. Arlington, VA.: Center for Applied Linguistics;

Haugen, E. (1973). Bilingualism, language contact, and immigrant languages in the United States: a research report, 1956-1970. In Thomas Sebeok (ed.), *Current trends in linguistics*. The Hague: Mouton;

Myers-Scotton, C. (1993). *Duelling languages: Grammatical structure in Codeswitching*. Oxford: Clarendon;

Pfaf, C. (1979). "Constraints on Language Mixing: Intrasentential Code-Switching and Borrowing in Spanish/English". *Language Related Research*. 55, 2;

Poplack, S. (1985). Aspects of multilingualism. *proceedings from fourth nordic symposium on bilingualism*

On some problematic Ossetic etymologies

SAMPSA HOLOPAINEN[1] & ANDRÁS CZENTNÁR[2]
[1]University of Helsinki, Finland; [2]ELTE University, Budapest, Hungary

Ossetic is one of the most thoroughly studied Iranian languages, being among the few that have an etymological dictionary (Abaev 1958–1989), and the bulk of the Ossetic vocabulary can be derived from Proto-Iranian (Thordarson 1989: 477). However, it is well-known that there are several words of uncertain origin in Ossetic, some words lacking convincing etymology altogether (Cheung 2002: 7). Some Ossetic words have been vaguely labeled as *Wanderwörter*, with uncertain relationship to words in other languages of the Caucasus, sometimes with uncertain relationship to words in Turkic languages, and in several cases to words in Hungarian or other Finno-Ugric languages.

In this presentation aims we show how knowledge of Turkic and Hungarian historical linguistics can enhance our knowledge of Ossetic etymology, and thus Iranian etymology in general, as well. A number of uncertain etymologies for Ossetic words will be critically discussed, with special attention paid to the assumed relationship of Turkic and Hungarian words. The main task is to define the direction of borrowing of these words, as well as to discuss the plausibility of Iranian explanations that have been suggested for some of these lexemes in earlier literature.

To cite some examples, Ossetic *kært* 'garden' has often been linked with Hungarian *kert* 'garden', and although the etymological connection of the two forms is obvious, it is complicated by the relationship to similar words in Turkic, such as Chuvash *karta*, and the irregular relationship of the Ossetic word to its assumed Iranian cognates (Róna-Tas & Berta 2011, s.v. *kert*); due to the *Wanderwort*-character of these words, this etymology deserves a fresh look. Ossetic *ʒedyr* / *ʒæduræ* 'blackberry' has a North-West Caucasian etymology according to Cheung (2017: 29), which is, however, problematic due to the assumed substitution of North-West Caucasian **z* by the Ossetic affricate and the hypothetical idea of a compound; relationship to Hungarian *szeder* 'blackberry' has remained unclear and often overlooked, even though a borrowing from Alanic to Hungarian has been suggested by Helimski (2002). Abaev (1965) mentions the outdated explanation of *ʒedyr* as a loanword from a Finno-Ugric source that is reflected by Mari *šaptə̑r* 'blackcurrant', Udmurt *suter* id., but we argue that this cannot be correct due to problems with Finno-Ugric phonology.

Other *Wanderwort*-like words in Ossetic that we will analyze include *pysyra* / *pursa* 'Urtica ures', which likewise has uncertain connections to similar words in Hungarian (*borsó* 'pea') and the Turkic languages (such as Karachay Balkar *murs*), recently discussed by Cheung (2017: 33), and *biræğ* / *beræğ* 'wolf', the relationship of which to Turkic *böri* 'wolf' and Finno-Ugric forms such as Mari *pire* 'wolf' remains unclear (Cheung 2002: 173).

References

Abaev, Vasilij I. 1958–1989: *Историко-этимологический словарь осетинского языка*. Vols. I–V. Moscow & Leningrad: Akademiya Nauk SSSR.

Abaev, Vasilij I. 1965: К алано-венгерским лексическим связям. – Gyula Orturay & Tibor Bodrogi (eds.), Bodrogi *Europa et Hungaria: Congressus Ethnographicus in Hungaria*. Budapest: Akadémiai Kiadó. 517–537.

Cheung, Johnny 2002: *Studies in the historical development of the Ossetic vocalism*. Beiträge zur Iranistik, Band 21; Wiesbaden: Dr. Ludwig Reichert Verlag.

Cheung, Johnny 2017: The Contacts between the Ossetians and Karachay-Balkars, According to V.I. Abaev and Marrian ideology. – Ramazan Korkmaz & Gürkan Doğan (eds.), *Endangered languages of the Caucasus*. Leiden: Brill. 17–38.

Helimski, Eugene 2002: Zu den iranischen Lehnwörtern im Ungarischen. – Rogier Blokland & Cornelius Hasselblatt (eds.), Finno-Ugrians and Indo-Europeans: Linguistic and Literary Contacts. Proceedings of the Symposium at the University of Groningen, November 22–24, 2001. Studia Fenno-Ugrica Groningana 2. Maastricht: Shaker Publishing. 105–111.

Róna-Tas, András & Árpád Berta (with the assistance of László Károly) 2011: *West Old Turkic*. *Turkic loanwords in Hungarian*. Wiesbaden: Harrassowitz:

Throdarson, Fridrik 1989: Ossetic. – Rüdiger Schmitt (ed.), *Compendium Linguarum Iranicarum*. Wiesbaden: Reichert. 456–479.

The Passive Voice in Modern Persian: An Alternative to Intransitive Function

SAYURI IGARASHI
International University of Health and Welfare, Tochigi Prefecture, JAPAN

1. Introduction

This presentation explains how the passive voice functions as an alternative expression in modern Persian for the missing intransitive function of transitive verbs that lack intransitive partners. In addition, the passive voice is indispensable for transitive verbs with their intransitive partners to clarify the difference in meaning between transitive and intransitive verbs. I shall discuss the above perspective that has not been discussed previously in the literature.

2. Data And Methods

extracted some intransitive and transitive verbs as well as past participles of transitive verbs used as adjectives in the Persian language from the works of Kuroyanagi (1996) and Dabīr-Moqaddam (1985).

egarding the relationship between intransitive/transitive verbs and passive sentences in Japanese, Nomura (1982: 169) explains that "aside from the group of transitive-intransitive verb pairs, there is a group of transitive verbs without their intransitive partners, and passive verbs complement the intransitive function that this type of transitive verb lacks." To clarify that the passive voice in Persian performs the same function, the transitive and intransitive verbs, adjectives, and transitive verbs in the passive voice extracted from the above works were observed in Japanese and in Persian. This revealed the complementary relationship between the passive voice and missing functions, as well as the transitive verbs that are rarely used as passive verbs.

3. Results

Result 1: The number of verbs and adjectives

Kuroyanagi (1996) lists 416 main verbs in Persian, of which 275 are transitive verbs, 179 are intransitive verbs, and 38 are verbs with both transitive and intransitive functions.[12]

Of the 275 transitive verbs, 179 do not have intransitive partners, while 96 have corresponding intransitive forms, and 89 of the 179 intransitive verbs are verbs without a transitive partner, while 90 have corresponding transitive forms. The list also includes 95 adjectives with the same form as past participles of transitive verbs forming the passive verbs.

Result 2: The relationship among transitive verbs, intransitive verbs, passive voice, and adjectives

he following table shows some of the extracted transitive and intransitive verbs, adjectives with the same form as past participles, and verbs in the passive voice in parallel:

Adjective	Passive	Intransitive	Transitive
ālūde (to be dirty)	ālūde šodan (to be made dirty)	ālūdan (to get dirty)	ālūdan (to dirty)
×	košte šodan (to be killed)	×	koštan (to kill)
oftāde (to be dropped)	×	oftādan (to drop)	×
×	×	×	būīdan (to smell)

[12] As the number of transitive and intransitive verbs includes the number of verbs that function as both transitive and intransitive verbs, the sum and the number of all the main verbs do not match.

4. Discussion

Discussion 1: The relationship among transitive verbs, intransitive verbs, passive voice, and adjectives

The passive form of the verb *ālūdan* (to get dirty / to dirty), that is, both transitive and intransitive, is used when an event is produced artificially (1a). When the adjective form is used as in (1b), emphasis is on the state rather than its cause.

（1a）*havā ālūde šode ast.* (The air is made dirty (polluted))
NOUN-air ADJ/Vt PASTP-to dirtyVi PRES-PF -to become

（1b）*havā ālūde ast.* (The air is dirty (polluted))
NOUN-air ADJ -to dirty Vi PRES-PF –to be

In the case of *koštan* (to kill), a transitive verb without its intransitive partner, the passive voice is used to describe the phenomenon of "being killed by someone" (2a). The motive for the use of the verb of the sentence is different from that of the transitive construction of (2b) in which the agent is clear. The fact that they need to be used to describe different events suggests the importance of the passive voice in Persian.

（2a）*mohammad tavvasot-e hamīd košte šod.* (Mohammad was killed by Hamid)
PSN-Mohammad PP-by …PSN-Hamid Vt PASTP-to kill Vi 3SG PAST-to become

（2b）*hamīd mohammad rā košt.* (Hamid killed Mohammad)
PSN-Hamid PSN-Mohammad POSTP-(indicates the object) Vt 3SG PAST-to kill

In addition, since the verb *oftādan* meaning "to drop" is intransitive and does not have the corresponding transitive form, it cannot be used in the passive voice. However, its past participle is used as an adjective to describe the results of both artificial actions and natural phenomena. The following is an example.

（3）*dar vasat -e īn ketāb yekī safhe oftāde dārad.*
PREP-in NOUN-between COMB-posessive D-ADJ-this NOUN-book NUM-1 NOUN-page ADJ-to drop Vt 3SG PRES-to have

(This book has one missing page.) (https://www.vajehyab.com/dehkhoda/)

Discussion 2: Verbs that are rarely used in the passive voice

Last, the observation revealed that many transitive verbs, such as *būīdan* (to smell), that describe one-time discontinuous actions are rarely used as passive verbs and do not have adjective forms. The passive voice in Persian is composed of the past participle of a transitive verb + *šodan* in principle, but since not all transitive verbs can be used in the passive voice, one must pay attention to the existence and characteristics of transitive verbs with which sentences in passive voice cannot be formed.

5. Conclusion

Earlier studies have focused on discussing the existence of the passive voice in Persian and explaining the syntax considered by researchers as passive. This study, however, discovered and demonstrated a new function of the passive voice: it complements the meaning of intransitive verbs absent in transitive verbs that do not have their intransitive partners. In other words, going beyond the description of previous studies conducted to confirm the existence of passive voice, this article argues that the passive voice plays a strong role in modern Persian, and thus, its existence is necessary.

References

Kuroyanagi, T. (1996) *Modern Persian Dictionary*, Tokyo: Daigaku Shorin.
Lazard, G. (1992) *A Grammar of Contemporary Persian*, San Diego: Mazda Publishers.
Moyne. J. A. (1974) The So-Called Passive in Persian, *Foundations of Language*, Volume 12, 249-267, Dordrecht: D. Reidel Publishing Company.
Nomura, T. (1982) "Intransitive, Transitive and Passive Verbs," *Japanese Language and Culture*, No. 11, 161-179, Department for International Research Students, Osaka University of Foreign Studies.
Phillott, D. C. (1919) *Higher Persian Grammar*, Calcutta: Calcutta University.
Windfuhr, G. L. (1975) *Persian Grammar. History and State of Its Study*, Hague: Mouton Publishers.
دبیرمقدم، محمد.(1985)مجهول در زبان فارسی، *مجله زبانشناسی*، ساله دوم شماره اول،31- 46،مرکزنشر دانشگاه.

Lexical Semantic Fields in Iranian Languages: The case of Bakhtiari Body Parts Terminology

SIMA ZOLFAGHARI
Leiden University

The study of different Iranian ethnic/ linguistic groups from an anthropological linguistic point of view has been sorely ignored in the field of Iranian linguistics. This is especially true in case of studies where various semantic domains are explored to open a window to the mind and the worldview of different peoples. As for the body parts terminology, there are a few valuable studies such as Filippone (2010) and (Abasi 2012), but a study which is based on a comparative cross-linguistic basis and studies the subject from an anthropological point of view was missing. The original research, Zolfaghari (2017) from which this talk is extracted was designed to investigate the Bakhtiari lexicon, in an attempt to understand a Bakhtiari nomad's cognitive picture of the world through his lexical reservoir and the way he classifies his natural environment.

The theoretical basis of this investigation has been summarized in the following questions from Majid (2006: 241): "How people categorize the world? This is one of the fundamental issues faced by researchers in linguistics, psychology, anthropology, and cognitive sciences. Is categorization the same between individuals, either as a result of innate concepts, or regularities in the perceptual array? Or, is human categorization arbitrary—a matter of cultural or linguistic convention?" The aim of that study was to '… to understand the relative impact of universal versus culture specific principles of categorization' (Majid, 2006: 242). When it comes to the body parts, the authors further express their specific questions 'concerning categorization and linguistic/conceptual segmentation of the body' as follows: "How do languages conventionally segment the body into parts? Does the set of body part terms constitute a structured system in all languages? Is there a universal, cross-linguistically consistent way of categorizing the body?" (Enfield et al. 2006: 138). The study of Bakhtiari body parts terminology will also draw on other important and relevant sources on lexical semantics, such as: Andersen (1978), Brown (1976) and Brown & Witkowski (1981). In regard to body parts, Brown (1976) designed a research plan based on the "naming behavior" of forty-one globally distributed languages and accordingly she drew "twelve general principles of classification and nomenclature in human anatomical partonomy". Andersen (1978), in turn, used Brown's data, added by a few more languages, calculated nine principles of categorization shared as universals in lexical fields.

These principles were tested with Bakhtiari data (306 terms in total), which revealed the fact that Bakhtiari segmentation of the body parts and also their nomenclature does not always conform with the proposed lexical universals, for instance, unlike many languages, Bakhtiari does not a have a word to refer to 'muscle' as a result of their different way of segmenting the body. Bakhtiari has a relatively abundant terminology of body parts. One manifestation of this characteristic is the existence of multiple words for the same part of the body, sometimes with no obvious difference in their meanings, registers or style, for instance: ***ri, šelg***, ***belešt, reχt*** and ***su/iræt*** , all to refer to 'face'. Another example would be the terms that refer to 'body': ***laš***, ***tæn***, ***bæðæn*** and ***ænḍo/am.*** Some of these similar terms are the result of borrowing from the dominant Persian language, but many others, plausibly, have been acquired through centuries of contact with neighboring languages and cultures within the borders of the present day Iran and beyond. The other reason for this diversity, in my opinion, lies in the special lifestyle of Bakhtiari nomads. Brown & Witkowski (1981: 207) talk about the 'intervening cultural variable' in polysemy development and loss: "While intrinsic perceptual salience is relatively constant across languages, cultural significance can vary greatly". As a pastoral, herd-breeder, it is crucial for a Bakhtiari nomad to be able to

define and explain different parts of his animals' body with precision. This is because for generations, they themselves were responsible for taking care of their animals in case of any accident, delivery or disease. Also in order to trade their animals they should be able to define all the physical characteristics of the livestock. This attention to animal body parts is a probable explanation for developing a detailed vocabulary to define different human limbs and organs, too. There is yet another probable historical reason for these multiple forms that harks back to the dominant pre-Islamic Zoroastrianism on the Iranian plateau. In Zoroastrian cosmology, the whole of creation was divided into two parties: Angra Mainyu "evil spirit/mind/thought" as opposed to Sepanta Mainyu "bounteous spirit" associated with Ahura Mazda, the Creator. Based on these ideas, in some cases, two sets of words and terms, even for body parts, were used to refer to things associated with these two opposing concepts. This distinction dissolved in the course of time, especially by the dominance of Islam, which vanquished the prevailing Zoroastrianism. Reminiscences of these pairs can be found now, used with almost identical meaning, in different Iranian languages such as Bakhtiari.

In this talk body parts terminology and their categorization by Bakhtiaris will be presented. The related data were extracted from Maddadi (1996) and also by eliciting data in actual and virtual fieldwork. The data is organized in six tables: 1) Face and its parts, 2) External parts of the body, 3) Internal parts of the body, 4) Bodily products, 5) Additional body parts and products of animals and 6) Configuration parts of the body. In each table the morphological aspects of the words are also brought into consideration, as well as some other cultural related concepts of the body parts such as their corporal merits and other associations. This latter cultural related part of the original study is developed to cover the concept of the 'Embodiment' as has been brought into attention by a number of scholars in recent years (Enfield & Wierzbicka: 2002; Sharifian, Dirven & Niemeier: 2008; E and Maalej & Yu 2011). The main focus of this part of the study is to explore how in the Bakhtiari culture, individual body parts are deployed to conceptualize an emotion, a cultural value or a human characteristic.

References

Abasi, B and A.Gheituri. 2012. Semantics of Hawrami Body part Terms. *Iran and the Caucasus 16* (2012) 319-335.

Andersen, E. S. 1978. Lexical Universals of Body-part Terminology. In, J. H. Greenberg (ed.), *Universals of Human Language*, vol.3.: 335-368. Standford University Press.

Brown, C. H. 1976. General principles of human anatomical partonomy and speculations on the growth of partonomic nomenclature. *American Ethnologist 3*: 400-424.

Brown, C. H., Witkowski, S.R. 1981. Figurative language in a universalist perspective. *American Ethnologist 8*: 596-615.

Enfield, N., & Wierzbicka,A. (Eds.).2002. Special issue on the body in description of emotions. *Pragmatics and Cognition, 10* (1/2).

Enfield, N. J. et al. 2006. Elicitation Guide on parts of the Body. *Language Sciences 28*: 148.

Filippone, Ela. 2010. The Fingers and their Names in the Iranian Languages (*Onomasiological Studies on Body-Part Terms, I*). Wien: Österreichische Akademie der Wissenschaften.

Maddadi, Zohrab. 1996/ 1375.[*Lexicon of the Bakhtiari language*]. Ahvāz: Ayat publications.

Majid, A. /M. v. Staden. 2006. Cross-linguistic Categorization of the Body. Introduction. *Language Sciences 28*: 137-147.

Majid, A. 2010. Words for parts of the body. In B. C. Malt, & P. Wolff (Eds.), *Words and the Mind: How words capture human experience*, New York: Oxford University Press: 58-71.

Sharifian, F., Dirven, R., Yu, N. & Niemeier, S. (Eds.). 2008. *Culture, body and language: Conceptualizations of internal body organs across cultures and languages*. Berlin and Vew York: Mouton de Gruyter.

Zolfaghari, Sima. 2007. *An Anthropological Survey on Bakhtiari Words in Comparison with Middle Persian*. M.A. thesis. Tehran University.

Zolfaghari, Sima. 2017. *The Bakhtiaris: An Anthropological-linguistic lexical study of Haft Lang nomads of southwest Iran*. Ph.D. Dissertation. Leiden University. Leiden.

Traces of Genitive Case in Southern Zazaki Ezafe

SONGÜL GÜNDOĞDU & ARSALAN KAHNEMUYIPOUR & ANDREW PETERS
University of Toronto

Ezafe is a common feature of many Iranian languages wherein a vowel is inserted between the head noun and post-nominal modifiers (including possessors) iteratively (Samiian 1994, Ghomeshi 1997, Samvelian 2008, Larson & Yamakido 2008, Haig 2011, a.o), as in the following schema: [NOUN-EZ1 MOD1-EZ2 MOD2]. The present study investigates the distribution of EZ2 in Southern Zazaki, an Iranian language spoken in Southeastern Turkey. The form of the Ezafe in Zazaki is *-a/-o/-ê* depending on the number, gender and case of the head noun. Meanwhile, in Southern Zazaki (in contrast with Northern and Central dialects), EZ2 sometimes appears as *-a/-o/-ê* and sometimes as *-dê/-da* (Todd 2002, Paul 1998, 2009, Keskin 2010, Werner 2018). Our goal is to account for the contexts in which *-dê/-da* appears. We propose that the Ezafe morpheme appears as *-dê/-da* when a modified noun [NOUN-EZ MOD] is placed in a context where it receives genitive case, either in a possessive construction or from a selecting adposition. All data herein comes from the authors' fieldwork with native Southern Zazaki speakers (from Siverek).

In Zazaki, the Ezafe morpheme inflects for number and gender of the head noun (1). In the possessive context, this variation appears as *-a* for feminine and *-ê* for plural or masculine (2).

(1) a. kutık-o sıpê
dog-EZ.M white
'the white dog'

b. bız-a sıpê
goat-EZ.F white
'the white goat'

c. bız-ê sıpê
goat-EZ.PL white
'the white goats'

(2) a. kutık-ê mın
dog-EZ.M my
'my dog'

b. bız-a mın
goat-EZ.F my
'my goat'

c. bız-ê mın
goat-EZ.PL my
'my goats'

Meanwhile, when the modified noun phrase appears in a possessive construction leading to two instances of Ezafe, the second Ezafe (EZ2) is realized as *-dê/-da,* with *–dê* used for masculine (or plural) and *–da* for feminine. In (3), we see this for the examples with adjectival modification and in (4) with recursive possessive constructions:

(3) a. goş-ê kutık-dê sıpê
ear-EZ.M dog-DE white
'the white dog's ear'

b. şıt-ê biz-da sıpê
milk-EZ.M goat-DA white
'the white goat's milk'

(4) a. goş-ê kutık-dê mın
ear-EZ.M dog-DE my
'my dog's ear'

b. şıt-ê biz-da mın
milk-EZ.M goat-DA my
'my goat's milk'

At first glance, it may appear that generally in the context of [NOUN-EZ1 X-EZ2 Y], EZ2 appears as *-dê/-da*. However, a closer examination reveals that EZ2 appears as *-dê/-da* only with the following phrasing [NOUN-EZ1 [X-EZ2 Y]] as in (5) (see also Paul 1998, 2009, Todd 2002), and not [[NOUN-EZ1 X]-EZ2 Y] as in (6):

(5) şıt-ê mangar-da siya
milk-EZ.M cow-DA black
'the black cow's milk'

(6) şıt-ê sıpê-yê manga
milk-EZ.M white-EZ.M cow
'the cow's white milk'

Meanwhile, EZ2 does not appear as *-dê/-da* in all [NOUN-EZ1 [X-EZ2 Y]] contexts either, as we see in the contrast between (5) and (7). In (5), [X-EZ2 Y] is in a possessive relation with the head noun, whereas in (7), [X-EZ2 Y] is a complex modifier in an adjectival relation with the head noun. This shows that *-dê/-da* does not appear in all contexts where a phrase containing Ezafe is embedded in a larger Ezafe construction (*à la* "doubled Ezafe" of Larson & Yamakido (2006) or "dependent Ezafe" of Todd 2002). Rather, the specific genitive case relation with the head noun is crucial.

(7) sol-ê siya-yê tari
shoe-EZ.PL black-EZ.PL dark ‘the dark black shoes’

In addition to the above contexts, we further note that *-dê/-da* also appears when [NOUN-EZ MOD] is selected by adpositions (8). We posit that adpositions assign genitive case to their complements in Zazaki, as found elsewhere (e.g. in Balochi postpositions, Jahani & Korn 2009).

[*-dê/-da* has also been observed in the context of allatives (i.e. postverbal goals) and indirect objects (Paul 1998, Werner 2018), an issue we will also discuss in our presentation.]

(8) a. mase-dê siyay sero
table-DE black on
‘on the black table’

b. war-da mın ra
sister-DA my from
‘from my sister’

Morphologically, the case marking on the complement of the adposition and the direct object of a present verb are the same in Zazaki, traditionally termed as “oblique”. However, a comparison between the realization of Ezafe in these two contexts reveals that these two cases need to be distinguished syntactically. In (9), we see that the Ezafe on the same modified noun as (8), used as the direct object of a present verb, does not appear as *-dê/-da* (see also Paul 1998, Todd 2002, Werner 2018). This is best illustrated by the contrast in (10) and (11). These examples establish that the presence of genitive case (as distinct from what has been termed as oblique) is key in accounting for the distribution of *-dê/-da*.

(9) Eza mase-yê siya vinenn-a
I table-EZ.M.OBL black see.PRS-1S ‘I see the black table.’

(10) Eza kutık-ê gırdi vinenn-a
I dog-EZ.M.OBL big see.PRS-1S ‘I see the big dog.’

(11) Eza kutık-dê gırdi re unen-a
I dog-DE big at look.PRS-1S ‘I look at the big dog.’

To summarize, we propose that *-dê/-da* is the result of an Ezafe construction receiving genitive case, either within a possessive noun phrase or by an adposition. In other words, *-dê/-da* is the morphological realization of *EZ.GEN* with sensitivity to gender/number. This is shown schematically in (12). (DE stands for both *-dê* and *–da.*)

(12) a. $_{NP}$[NOUN-EZ1 $_{NP\text{-}GEN}$ [X-DE Y]] b. $_{PP}$[$_{NP}$[NOUN-DE X] P_{GEN}]

In this study, we have provided an account for the distribution of *-dê/-da* in Southern Zazaki. This account crucially relies on the presence of genitive case in the language, as *-dê/-da* is restricted to Ezafe being embedded in a genitive context, either possessive or adpositional. This shows that while genitive case is not morphologically realized elsewhere in the language, its traces can still be found with morphosyntactic consequences.

References:

Ghomeshi, J. 1997. Non-projecting nouns and the ezafe: construction in Persian. *NLLT*, *15*(4), 729-788. **Haig, G. 2011.** Linker, relativizer, nominalizer, tense-particle: On the Ezafe in West Iranian. In *Nominalization in Asian Languages: Diachronic and Typological Perspectives*, 363-390. **Jahani, C. & Korn, A. 2009.** Balochi. In *The Iranian Languages*, 634-692. **Keskin, M. 2010.** Zazaca üzerine notlar. *Herkesin Bildiği Sır: Dersim*, 221-244. **Larson, R. & Yamakido, H. 2006**. Zazaki “Double Ezafe” as Double Case-marking. Paper presented at the LSA Meeting. **Larson, R. & Yamakido, H. 2008.** Ezafe and the Deep Position of Nominal Modifiers. In *Adjectives and Adverbs: Syntax, Semantics and Discourse.* 43-70. **Paul, L. (1998).** Zazaki. Grammatik und versuch einer Dialektologie, Wiesbaden. **Paul, L. 2009.** Zazaki. In *The Iranian Languages*, 545-586. **Samiian, V. 1994.** The Ezafe construction: Some implications for the theory of X-bar syntax. *Persian Studies in North America*, 17-41. **Samvelian, P. 2007.** A (phrasal) affix analysis of the Persian Ezafe. *J. Linguistics*, *43*(3), 605-645. **Todd, T. L. 2002.** *A grammar of Dimili: Also known as Zaza*. **Werner, B. (2018).** Forms and Meanings of the Ezafe in Zazaki”: In Endangered Iranian Languages.

Building a Tagged Corpus of the Classical Persian Language

TATIANA KONONOVA
Institute of Asian and African Studies, Moscow State University, Moscow, Russia

A tagged corpus is an invaluable instrument for linguistic research. This work was dedicated to building such a corpus for the Classical Persian language. The Classical Persian is understood as the New Persian language used in the medieval times and represented mostly by literary texts. The term is opposed to the Contemporary Persian that is spoken today. Despite many similarities between the two, the classical language is characterized by many peculiarities partly described in [Rubinchik, 2001], [Lazard, 1975], [Paul, 2008] and others.
An open-source collection of the Classical Persian poetry (IX-XVII centuries) was downloaded from an Internet library [8] and as a result a 4.3 million token corpus with 120000 unique word forms was collected.
Further corpus processing included tokenization, lemmatization and morphological and part-of-speech tagging. It was realized in several steps.

1. Tag-set

A set of 10 part-of-speech tags with corresponding morphological tags (or attributes) was developed. It was based on contemporary Persian tag-sets: most of Peykare corpus tag-set **[**Bijankhan et al., 2011] combined with some attributes from [QasemiZadeh et al., 2006]. An attempt was made to capture features specific for the classical language. Most importantly all fused forms were considered as one token and corresponding tags were introduced to reflect that (e.g. pronominal enclitics, prepositions and conjunctions that can be fused with most parts of speech). The resulting system was meant to provide a detailed morphological description of graphical tokens, for example:
بدستانت *be-dast-ān-at* N PREP PL PRO2SG “to your hands” (with lemma *dast* “hand”),
شادتریم *šād-tar-im* AJ COMP COP1PL “we are happier” (شاد *šād* “happy”),
آوریمش *āvar-im-eš* V IND PRES 1PL PRO3SG «we bring him» (آوردن *āvardan* «bring»).

2. Rule based lemmatization and morphological and POS tagging

On the next stage a lemma and tags from the designed tag-set were automatically assigned to each word form (including fused ones) of the collected corpus. The process was rule-based.
For that a list of 58872 words with corresponding POS tags was automatically extracted from the digital version of Persian-Russian dictionary [Rubinchik, 1970]. Besides, a list of 2465 irregular plurals for nouns and 986 present stems for verbs was also collected from the same source. These word lists were used to generate word forms that were matched against the corpus. The number of theoretically possible forms for every dictionary word could reach millions, because generation rules included orthographical variations, fused prepositions, pronouns and conjunctions etc. For example, one of the rules for noun forms generation looks as following: *conjun.* + *preposition* + *stem* + *plural* + *pronoun* + *article* + *copula*+ *conj.ke.*
Lemmatization and tagging were realized through matching the generated forms against the corpus. This system allowed to tag 75% of 120 000 unique word forms that accounted for 95.7% of 4.3 million tokens. However, the system’s ambiguity level was high: a word could get several independent interpretations (including various parts of speech tags) and their average number amounted to 4.4. That was corrected on the next stage.

3. Statistical part-of-speech tagging

To reduce ambiguity level rule based POS tagging was paired with statistical one. A number of POS-taggers trained on Contemporary Persian corpora were tested. The best result (92.4% accuracy for a 1000 word sample corpus) was achieved by Stanford-tagger trained on

Peykare corpus from Hazm library [7]. The output had to be adapted because of some differences in tag-sets. Combining the results of both taggers significantly lowered ambiguity level from 4.4 interpretation per word to 2.2 as shown in Table1.

Table 1. Ambiguity level with and without statistical POS tagging

Average number of	Rule based tagging and lemmatization	Rule based + statistical POS-tagging
Interpretations per word	4.4	2.2
Lemmas per word	2.1	1.4
POS tags per word	1.7	1

4. Quality assessment

The resulting tagging quality of the hybrid two-stage system was measured on a 1000 word sample subcorpus (see Table 2).

Table 2. Tagging quality assessment

	Corpus		Sample subcorpus	
Tokens	4 346 493	100%	1000	100%
Correct POS tag	n/a	n/a	924	**92.4%**
Rule based tagger assigned a tag	4 159 594	95.7%	955	95.5%
Rule based tagger failed to assign a tag	186 899	4.3%	45	4.5%
POS assigned by statistical tagger coincides with one of the tags of rule based tagger	4 307 374	99.1%	993	99.3%
Correct POS tag and at least one of the morphological interpretations is correct	n/a	n/a	918	**91.8%**

Quality assessment demonstrated high accuracy level. For 92.4% of words POS tags were accurate. Although tagging ambiguity was still present (on average a word received 2.2 morphological interpretations and 1.4 lemmas) in 91.8% of cases one of the assigned interpretations was also correct. That gives a reason to consider the developed system satisfactory.

References:

1. Bijankhan Mahmood, Sheykhzadegan Javad, Bahrani Mohammad, Ghayoomi Masood. Lessons from building a Persian written corpus: Peykare (2011). Language Resources and Evaluation. Vol. 45, No. 2 (Spring 2011), pp. 143-164.
2. Lazard G. “The Rise of the New Persian Language,” in R. N. Frye, ed., The Cambridge History of Iran IV. The Period from the Arab Invasion to the Saljuqs, Cambridge, 1975, pp. 595-632
3. Paul L., “Remarks on the Evolution of the New Persian System of Past Tenses,” in S. Shaked and A. Netzer, eds., Irano-Judaica. Studies Relating to Jewish Contacts with Persian Culture throughout the Ages VI, Jerusalem, 2008, pp. 185-95.
4. Rubinchik Yu. Grammatika sovremennogo persidskogo yazyka [Grammar of the contemporary Persian language]. Moscow, “Vostochnaya literature” RAN, 2001. 5. Rubinchik Yu. A. Persidsko-russkiy slovar [Persian-Russian dictionary]. Moscow, Russkiy yazyk, 1970.
6. QasemiZadeh Behrang, Rahimi Saeed. (2006) Persian in MULTEXT-East Framework. Proceeding FinTAL'06 Proceedings of the 5th international conference on Advances in Natural Language Processing, pp. 541-551.
7. Internet source: https://github.com/sobhe/hazm
8. Internet source: www.rira.ir