Masoud Mohammadirad* and Pollet Samvelian Procliticization as a residual of second positioning: evidence from West Iranian languages

https://doi.org/10.1515/flin-2021-2024 Received February 14, 2020; accepted April 12, 2021

Abstract: Person clitics show proclitic attachment in some West Iranian languages. Nevertheless, most of the literature has continued to focus on enclitics. This paper provides evidence that a good number of modern languages have developed proclitics, presumably from the middle Iranian period onward. Using synchronic data from modern languages gathered in the field, and contrasting it with the Middle Iranian period and current clausal second-position clitic systems, we develop some hypotheses regarding the rise of proclitics in modern languages. We argue that proclitic attachment has resulted from the reanalysis and/or the loss of clause-initial clitic hosting particles of the Middle Iranian period, and the actualization of the stray clitic as a proclitic on some host to the right. This trajectory from second position enclitics to proclitics, which is also attested in Old Romance and Uto-Aztecan, is argued to have been triggered by head attraction and rightward drift of clitics from clause-second position toward the verb in modern languages, giving rise to VP-based and Verb-based cliticization systems.

Keywords: actualization; person clitic; reanalysis; rightward drift; second position clitics

1 Introduction

Iranian languages constitute one of the branches of Indo-European language family. The oldest stages of Iranian languages are attested in Gatha Avestan, which is closely related to the earliest attested forms of Indo-Aryan, namely Vedic. In addition, Old Iranian is also attested in Old Persian texts, which are datable back to 500 BCE. Middle Iranian (beginning in the third century BCE), and New Iranian

^{*}Corresponding author: Masoud Mohammadirad, University of Sorbonne Nouvelle – Paris 3, Paris, France, E-mail: masoud.mohammadirad@sorbonne-nouvelle.fr. https://orcid.org/0000-0002-8531-5524

Pollet Samvelian, Lattice UMR 8094, University of Sorbonne Nouvelle – Paris 3, Paris, France, E-mail: pollet.samvelian@sorbonne-nouvelle.fr. https://orcid.org/0000-0002-6622-4117

(beginning in around the seventh century CE) are other stages of Iranian languages (Windfuhr 2009: 5).

Traditionally, Iranian languages are classified into two main groups of Eastern and Western sub-branches, each with their own subgroupings based on Northern and Southern poles: thus, for example, the Western branch is subdivided into Northwest and Southwest groups (Schmitt 1989; Windfuhr 2009). This classification is not clear-cut, that is, languages show gradience on the scale of being Southwestern versus North-western (see Paul 1998). More recently, Korn (2016, 2019) proposes to add a novel 'Central Iranian' sub-branch to the bipartite classification. In this paper we continue to use the term 'West Iranian' for the languages investigated. To avoid the confusion arising out of the tripartite classification of Iranian, we propose a classification of languages into micro-area 'language groups', as seen in Figure 1. These groupings roughly correspond to the existing classifications of Iranian languages in Schmitt (1989) and Windfuhr (2009), and tend to show



Figure 1: West Iranian languages investigated here.

more clearly the phenomenon under investigation here. That is, proclitics are a feature of Central Plateau languages, languages of Southeast Iran, and Nowdani in the Southwest language group (see Figure 3).

The following typological features characterize most modern Iranian languages (Windfuhr 2009: 31–34): tense-split ergativity, restricted to past tense verb forms derived from verbal participles, and differential object marking. Haig (2017: 467) adds other typological features: OV word order, and a very high frequency of complex predicates, based on a small set of light verbs.

In addition, the majority of West Iranian languages employ person clitics (or clitic pronouns). The latter are one of the most complex features of Iranian syntax, and have been subject to theoretical, typological, and diachronic studies (see Haig 2008, 2018; Jügel and Samvelian 2016; Korn 2009; Öpengin 2013; Samvelian 2007b, among others).

The clitic pronouns of West Iranian are reflexes of accusative and genitive/ dative clitic paradigms in Old Iranian (cf. Korn 2009). The clitic status of these forms in modern languages comes principally from (i) a low degree of selection and freedom of host selection; (ii) morphological idiosyncrasies in combination with hosts (see Section 3.2 for relevant examples).

Despite the vast literature on Iranian person clitics, some aspects of clitichood have nevertheless been overlooked in the previous scholarship. One such issue is the rise of proclitics in a subset of modern languages and their distribution. A survey of the literature suggests that apart from a few exceptions, e.g. Dabir-Moghaddam (2008), the tendency has been to consider encliticization as the sole means of clitic attachment across West Iranian. For instance, Korn (2009: 159) reports that "they [person clitics here] are used as enclitic counterparts of the stressed personal pronouns in all oblique functions". In the same way, Lecoq (2002: 86) regards the person clitics of Central Plateau languages as "enclitics". The underlying assumption behind these views is that in continuity with the Old Iranian and Middle West Iranian (henceforth MWI) periods, encliticization is the sole means of clitic attachment in modern West Iranian languages (henceforth WILs).

However, in a good number of WILs proclitic attachment is the only mode of adjoining for some person clitic functions – e.g. obligatory indexing of past transitive subjects (or A-past) and pronominal realization (or alternating indexing) of objects –, and an optional realization for others – e.g. prepositional complements (cf. Table 3). In those languages which have developed them, proclitics usually endorse the same set of functions as enclitics, including A-past NP, cf. (1), direct object, cf. (2), non-canonical subject, cf. (3), prepositional complement, cf. (4), and possessor, cf. (5) (see Section 4 for the extent of procliticization).

- (1) Nowdani
 malum-e ke bača-yl eš=xard-e¹
 obvious-cop.3sg COMPL child-PL 3sg:A-eat.PST-PRF
 'It is obvious that he has eaten the (goat)kids.'
 (SM. 34)
- Bastaki
 š=a-rest-e peš-e āsiābān
 3sg:0=IND-send.PRS-DRCT to-EZ² miller
 'He sends him to the miller.'
 (RS. 18)
- (3) Bandari *om=na-hasta*1sg:NC=NEG-exist.PRS
 'I don't have it.' [lit. 'to me it does not exist']
 (EL. 41)
- (4) Lari
 š=az_bar süt a-zan-en
 3sg:R=for whistle IND-hit.PRS-3PL
 'They whistle for him.'
 (PS1. 20)
- (5)Yazdi Zoroastrian уa mošta ārt e-kuz-ā $\check{s}=e$ gal punch flour IND-hit.prs-3sg foot one 3sg:poss=to '(The wolf) pours a handful of flour into his paw.' (SM2. 15)

¹ The function of each clitic is separated by a colon in the gloss after its person and number, hence, A: subject of past transitive verb; o: object; NC: non-canonical subject; R: adpositional complement or non-flagged secondary object; Poss: possessor. For practicality reasons, and for ease of understanding, the equals sign '=' is reserved for the glossing of person clitics and additive clitics. Other grammatical elements, e.g. the copula and the ezafe (see note 3), could also be said to have clitic status based on some clitichood criteria, e.g. promiscuous attachment. They are nonetheless separated by the hyphen '-'.

² The ezafe (from Arabic *`iḍāfa* 'joining, addition') is a head-marking linker used in the structure of the noun phrase in the majority of West Iranian languages. It links post-head modifiers to the head noun.

This paper aims to provide a first systematic survey of proclitic attachment in WILs. In addition, we address the rise of proclitics in WILs in the light of general typological tendencies. The data come from a corpus of natural speech and controlled speech (also including elicited data) developed in the form of a database of Iranian languages in Mohammadirad (2020). The examples are cited from the corpus along with the reference to their sentence number in a given text, hence 'Dashti, ZK. 36' means sentence number 36 from the text 'ZK' of Dashti. The elicited data are marked by 'EL'. Many of the languages studied here are highly endangered and have not been studied as yet. Some do not have ISO identifiers, nor are they indexed in Glottolog (see the Appendix).

In what follows, we will first give an overview of clitic placement in MWI. Section 3 provides a synopsis of cliticization systems in modern WILs. In Section 4, the extent of proclitic attachment across modern languages will be covered. Section 5 reviews the existing accounts of the rise of proclitics. Section 6 relates the rise of proclitics to the reanalysis of a set of particles in clause-initial position, and the consequent actualization of the stray clitic as a proclitic on the immediate element to the right. Building on the similarities between second position enclitics, and clause-initial proclitics, Section 7 concludes that the latter are derived from the former. This shift from en- to pro-clitic is argued to be motivated by the rightward drift of clitics and their attraction to the heads.

2 Person clitics in Middle West Iranian

The known MWI languages include Middle Persian and Parthian. Clitic pronouns, already available in Old Iranian and displaying two sets of paradigms, i.e. genitive/ dative and accusative, collapsed into a single non-nominative, or general oblique, form by Middle Iranian (cf. Table 1; Korn 2009: 160). This one set of clitics acquired all the non-nominative functions of its ancestors, including possessor, direct object, indirect participant, and a bunch of non-canonical subjects (see Haig 2008: Ch. 2 for discussion). The resulting clitic paradigm of MWI has survived in many modern languages, though it is absent in some of them, e.g. Sangesari, Kurmanji Kurdish, Zazaki (Windfuhr 1975: 462; see Jügel and Samvelian 2016; Öpengin and Mohammadirad forthcoming,³ for a recent discussion).

MWI person clitics are second position clitics in the sense of Wackernagel: they attach to the first element of the clause. This is shown in the following examples,

³ Despite the general consensus on the disappearance of person clitics in Kurmanji (or Northern Kurdish) dialects, the authors bring new evidence that person clitics have survived in some Kurmanji dialects bordering the Central Kurdish speech zone.

		Middle Persian	Parthian
SG	1	-(u)m	
	2	-(u)t, -(u)d	
	3	-(i)š	
PL	1	<i>-n</i> (rare), <i>-mān</i>	-mān
	2	-(i)tān, -(i)dān	-tān
	3	-(i)šān	

 Table 1: Manichean Middle Persian and Parthian pronominal clitics (Korn 2009: 160).

where diverse syntactic categories host the person clitic, including the subject NP, cf. (6), the subordinator, cf. (7), the coordinating conjunction, cf. (8), the adverb, cf. (9)–(10), and the last element of the preceding clause, cf. (11):

- (6) Parthian tw=m'n 'yy xwd'y 2sg=1pL:poss cop.2sg lord 'You are our lord.' (Brunner 1977: 102)
- (7) Middle Persian
 d'=š'n prg'r rs'd
 till=3PL:POSS victory arrive.PRS.3SG
 'Until their victory will arrive.'
 (Brunner 1977: 105)
- (8) Early New Persian AP⁴=t nm'd-ym CONJ=2SG:R show.PRS-1PL 'And (we) will show you.' (Heston 1976: 90)
- (9) Parthian *čīd=mān* pāyēd
 always=1PL:0 protect.PRS.3SG
 '(It) will always protects us.'
 (Durkin-Meisterenst 2006: 75, cited in Haig 2008: 115)

⁴ Following the tradition in Iranian linguistics, the words of Aramaic origin, commonly called ideograms, appear in capital letters.

- (10) Parthian qdy=s'n dyd 'yy when=3pL:A see.pst cop.2sg 'When they saw you.' (Brunner 1977: 114)
- Middle Persian
 nm'c br'nd=š gwynd
 praying offer.prs.3pL=3sg:r say.prs.3pL
 'They offer worship; they will say to him.'
 (Brunner 1977: 114)⁵

The domain of cliticization is the clause in the above examples. In (11) the clitic is phonologically attached to the verb of the preceding clause but is syntactically part of the following clause. This occurs when the clitic cannot be hosted by an appropriate host in its local clause. In (11), the clause in which the clitic occurs is comprised of the verb. Since the verb is not usually an anchoring element, the clitic is attached to the immediately-preceding element, which is the verb *br'nd* in the preceding clause.

Among the hosts for clitics, clausal conjunctions are of special interest to us. One such conjunction is the 'and-coordinator' *-ud*, which, together with its *sandhi* form *u*-, primarily had the role of connecting words, phrases, and clauses in WMI (Brunner 1977):

(12) Parthian

wcn **'wt** *zyrd pdmwxtn cy* '*dy*'wr'n voice *conj* heart grinding *ez* friend.*pl.obl* 'The grinding on of the voices and hearts of friends.' (Brunner 1977: 226)

The *sandhi* form *u*- also marked the beginning of the sentence, in that it acted as a clause-initial particle to which the clitic pronouns could attach.⁶ This is especially relevant in Middle Persian Pahlavi texts (Brunner 1977: 227).

(13) Parthian $u=\check{s}$ gurg $\bar{e}w$ grift PTCL=3sG:A wolf one catch.PST 'He caught a wolf.' (Durkin-Meisterenst 2014: 311)

⁵ See Brunner (1977: 114) for explanation on the reading of this sentence.

⁶ Brunner (1977: 227) refers to this use of u- as "quasi-adverbial", while Heston (1976: 249) uses the term "clause-marker".

(14) Middle Persian

u=tfrabīh-tarkardhēmPTCL=2sG:Afat-CMPRdo.PSTCOP.1SG'You made me fatter.'(Durkin-Meisterenst 2014: 432, mpB. 961)

(15) Parthian u=**šān** andar šhr hw cvš pār 'st ny PTCL=3PL:NC in DEM world thing debt exist.prs NEG '[...] They have no debts at all in this world.' (Brunner 1977: 228)

In these examples, the coordinating meaning of u- is less clear. The u- functions rather as a particle that assures the clausal second positioning of person clitics, hence our use of the term S2-assuring particles. Note further that the conditioning factor for u- to be a clitic hosting element is seemingly the absence of other eligible clitic hosts, such as the subject NP, and clausal adverbs, cf. (13)–(15).

Likewise, Jügel (2017) implies that the particle u- acts as a clitic host in Middle Iranian: "Enclitic pronouns frequently attach to the conjunction ud 'and', which then takes the form u-. This combination is so common that u- is also used when the meaning 'and' is not intended, i.e. u- becomes a semantically empty carrier for the enclitic pronoun".

The development of another S2-assuring particle is also of importance. This particle is derived from the adverb \bar{a} (') 'thus, then' (Brunner 1977; MacKenzie 1971). Brunner (1977) suggests that the sandhi form of the adverb ah, i.e. a-, might be a source for such a particle in Middle Persian Pahlavi texts.

(16) Middle Persian

a=*t tl mynyt* PTCL=2SG:0 LVC think.PRS.3SG 'He scorns you.' (Brunner 1977: 114)

(17) Middle Persian

 \bar{a} =m fr $\bar{a}z$ guft h \bar{e} , zarduxšt PTCL=1SG:A to say.PST COP.2SG Zoroaster 'I said to you, Zoroaster.' (Durkin-Meisterenst 2014: 398, mpB. 798)

A reflex of this particle has remained in the clitic paradigms of some Central Plateau languages:

- (18) Delijani
 āw ašon=a-bar-a
 water 3PL:0=IND-take.PRS-3SG
 'The water will displace them.'
 (GX. 18)
- (19) Khansari šomā ež=e-vin-di
 2PL 3sg:0=IND-see.PRS-2PL
 'You see him.'
 (QB. 17)

A final point to consider is that although clausal second positioning was the regular placement for person clitics in MWI, there is nevertheless some degree of flexibility in the original second-positioning rule for clitic placement. For instance, in (20) the prepositional complement clitic is attached to its governing head and not to the relative marker in the S2 position, marked by the 'underscore'. Likewise, in (21) the A-past clitic has skipped both the subject NP and the relative marker to appear on the prepositional phrase:

(20)Middle Persian andar šab ō wiyābān-ēw pad=**iš** ēč mad, kē in night desert-INDF nothing to came REL. in=3sg:R āb ud xwarišn пē būd water CONI food NEG exist.pst 'At night, he got in one desert where there was no water and food.' (Durkin-Meisterenst 2014: 402)

(21) Parthian

xrd_cyd_'cbw=tpdgtyftwisdomRELfromBuddha=2sg:Atook'The wisdom which you received fromBuddha.'(Brunner 1977: 102)

Later stages of Iranian provide ample evidence for an overall rightward drift in clitic placement rules, leading to abandonment of the second-positioning rule for the majority of languages⁷ (though it was retained in a minority, see Section 3.1): the relevant host for clitics in the modern languages is now some constituent of the VP, which may include the verb itself (see Sections 3.2 and 3.3). This move meant

⁷ Haig (2008: 334–338) argues that 'head attraction' and rightward drift were accountable for the changing rule of clitic positioning in Iranian languages.

that the necessity to maintain clitic-assuring particles relaxed, and facilitated their being re-analysed in some languages. However, the retention of S2-assuring particles is not assumed to precede historically the rightward drift of clitics: that is, as illustrated by (20)-(21), a language may maintain S2-assuring particles while at the same time having undergone rightward drift for some clitic functions, most notably possessors and preposition complements.

3 A synopsis of cliticization in West Iranian languages

Most modern WILs have abandoned the clause as the domain of cliticization in favour of VP-based and V-based cliticization domains (cf. Mohammadirad 2020:



Figure 2: Cliticization domains in West Iranian languages.

Ch. 5). This fact is illustrated in Figure 2. In the following subsections we give an overview of each of these cliticization domains.⁸

3.1 The clause as the domain of cliticization

As shown in Figure 2, Davani, Dashti, and Behbahani are languages which have preserved the older clausal second positioning of person clitics. As in MWI, person clitics in their different functions are hosted by a variety of clause-initial elements, including the subject NP, cf. (22), the clausal conjunction, cf. (23), clausal adverbs, cf. (24)–(25), and the clausal topic, cf. (26).

(22)	Behbahani sang= ey ser-e gerdu eškeni stone=3sg:A head-Ez walnut break.PST 'The stone broke walnut's head.' (SG1. 2)	
(23)	Dashti $t\bar{a}=t$ moraxas \emptyset - $\bar{a}i$ that=2sg:0 released IRR-give.prs.3sg 'That he let you go.' (ZK. 36)	
(24)	Davani <i>diar=šu čaqu keš-e-se</i> already=3 _{PL:A} knife pull. _{PST-PTCP-COP.3_{SG} 'They have already pulled out (the) knife.' (KS. 35)}	
(25)	Davani $\tilde{s}ad=\tilde{s}$ a $del-e_{-}$ da maybe=3sg:poss from heart-DEF PV 'Maybe he can soothe him.' [lit. 'pull (it) out (XX. 39)	ar-bār-e B-bring.prs-3sg from his heart']

⁸ Our stance in this article is that prosody is not significant for the positioning of clitics – though historically it may have been a relevant factor for clitic placement. In fact, most of the literature on Iranian clitics regards their placement as being determined either by morphology, or syntax (cf. Dabir-Moghaddam 2008; Haig 2008, 2013; Karimi 2010; Samvelian 2007a, 2007b, among others). An exception is the syntactic-prosodic analysis of clitic placement in Central Kurdish (Öpengin 2013).

(26) Davani $ma = \check{s}$ tā kasi das=om na-bas-se aso hand=1sg:poss 1sg=3sg:Atill now somebody NEG-tie.pst-prf 'I – nobody has chained me (my hands) yet.' (Mahamedi 1982: 454)

Among modern languages with the clause as the cliticization domain, Dashti and Davani are more conservative in having preserved certain features of S2positioning seen in MWI: first, in both languages the anchor can extend to the last element of the preceding clause. This is illustrated by the placement of A-past clitics below:

(27) Dashti veho mi-bi-e=**t** mi-košt to рā то 2sg foot IPFV-become.PST-2SG=2sG:A IPFV-kill.PST suddenly 1sg 'All of a sudden, you would get up (and) kill me.' (KX. 9) (28) Davani o=mu nana-i $hi=\mathbf{\check{s}}$ ya grandma-INDF PTCL=1PL:NC one exist.pst=3sg:A teli mi-ke doros round.bread right IPFV-do.PST 'We had a grandma who would bake bread.' (XX. 2)

Second, both languages have preserved a reflex of the S2-assuring particle *u*- of the Middle Iranian period, in contrast to Behbahani. As in MWI, the *u*- particle acts as an anchoring element in the absence of other eligible clitic hosts, i.e. subject NP, clausal conjunctions, clausal adverbs.

(29) Davani

o=*t ya memuni hā-de* PTCL=2SG:R one party PVB-give.PRS.1SG 'That I throw a party for you.' [lit. 'that I give you a party'] (XX. 14)

(30) Dashti

o=**mu** ri xar mi-nā PTCL=1PL:A on donkey IPFV-put.PST 'We would put (the sack) on donkeys.' (ZK. 20) (31) Davani $0=\check{s}$ baček-e bad те-ата i child-DEM1 PTCL=3SG:NC bad IPFV-come.pst.3sg.Drct DEM 'She hated this kid.' (KS. 8) (32) Dashti mi-košt-an 0=**mu**

```
PTCL=1PL:A IPFV-kill.PST-3PL:0
'We would kill them.'
(EJ. 20)
```

The first constituent within the clause is the object NP in (29), the prepositional phrase in (30), the non-verbal element of the complex predicate in (31), and the verb in (32). By attaching to the particle *o*-, the second position clitics in (29)–(32) avoid taking complex predicates and non-subject arguments of the verb as their anchoring elements, hence their placement remains clause-second. Recall that clitic-hosting particles can coexist with the rightward drift and/or head attachment of clitics. In the following example, the S2-assuring particle holds the A-past clitic in the clause-second position while at the same time the prepositional complement clitic remains attached to its head:

(33) Dashti

o=**m** $b\bar{a}=\check{s}$ harf mi-ze PTCL=1SG:A with=3SG:R talk.PST IPFV-hit.PST 'I would talk with him.' (field note)

3.2 The VP as the domain of cliticization

As shown in Figure 2, the largest number of WILs opt for what roughly corresponds to the verb phrase (VP) as the domain of cliticization. Notice that VP is used here in a loose sense, roughly corresponding to the verb or the verbal complex (verb plus auxiliary, complex predicates, particle verbs, etc.) plus the direct object and possibly the oblique (or prepositional) object.⁹ This means above all that unlike in

⁹ It is very common in syntactic studies within formal (or even typological) frameworks to assume that direct and indirect (oblique or prepositional) objects are part of the VP constituent (or, in other words, the verb projection). This is also what has been assumed in various studies on the placement of clitics in Central Kurdish (Haig 2008; Karimi 2010; Öpengin 2013; Samvelian 2007a, 2007b,

S2-clitic systems, the subject NP, clausal adjuncts, conjunctions, and topics are not eligible to be clitic hosts: the clitic positioning is rather determined with respect to the first available phrase within the VP.

(34) Southern Central Kurdish šaw-ē_ kuř-akān=ī bāng kird night-INDF boy-DEF.PL=3sG:A call do.PST 'One night he called his sons.' (SB. 3)

(35) Delijani
mā=š_ nun=eš ba-pet
mother=3sg:poss bread=3sg:A PFV-bake.pst
'His mother baked bread.'
(GX. 6)

In (34) and (35), the clausal adjunct and the subject NP respectively are skipped for the placement of A-past clitics. The A-past clitic rather attaches to the first element of the VP, which is the object NP in (34), and the non-verbal component of the complex predicate in (35).

A major isogloss divides the VP-based cliticization systems into those which allow morphological elements, e.g. TAM prefixes and the negative formative, to be clitic hosts, and those which do not. In the first group of languages, person clitics intervene between the TAM prefixes and the verb stem, exhibiting a kind of endoclitic behaviour,¹⁰ cf. (36)-(37).¹¹

bin-on
:0-see.prs-1sg
j.

among others). Our aim in this paper is not to define the notion of VP or provide some diagnostics for its identification (which could be controversial). The sense in which we use the term VP is a very common one, i.e. the VP consists of the verb and all its non-subject dependents. Another way of defining the VP-placement of clitics would be to say that they attach to the earliest non-subject constituent of the clause (Karimi 2010: 699). Note however that examples (34) and (36), which include clause-initial adverbs, seem to fall outside of Karimi's conception of the VP.

¹⁰ Cf. Samvelian (2007a, 2007b) for an endoclitic treatment of clitics in Central Kurdish in contexts where the verb is the last resort for cliticization.

¹¹ These cases of endoclitic realization run against one of the important diagnostics of clitichood, i.e. as syntactic items clitics are not expected to interrupt morphological words (cf. Halpern 1998; Nevis 2000).

(37) Baneh Central Kurdish
 a=t-bird-īn bo šahr-ī bāzī
 IPFV=2sg:A-take.PST-1PL:0 to city-Ez game
 'You would take us to the amusement park.'
 (EL. 41)

In the second group, the anchor can only be a syntactic element, hence the unavailability of morphological prefixes for clitic hosting:

(38) Sivandi vaqti_ke det-e eyāl-ā me-word=eš /*me=š-word when girl-DEF child-DOM IPFV-bring.PST=3SG:A 'When the girl would give birth to the child ...' (HT. 7)¹²

(39) Luri
bāyad ma ba-ir-am=aš /*ba=š-ir-am
AUX 1sG IRR-grab.PRS-1sG=3sG:0
'I must catch it.'
(Amān Allāhī and Thackston 1986: 145)

3.3 The verb as the domain of cliticization

A third group of languages in Figure 2 are those which take the verb as the domain of cliticization, hence V-based clitic systems. It is primarily in this group of languages that the rise of procliticization is attested. It is also here that one can observe that the S2-assuring particles of MWI have undergone a whole set of shifts. Therefore, V-based cliticization systems are our primary concern in this paper.

Except for Semnani, which has enclitic attachment, cf. (40), all other V-based clitic systems have proclitic attachment. In any case, the consequence of V-based cliticization is the development of an affix-like behaviour for clitics, in the sense that they lose their mobility and become selective with respect to their host. In the following examples, the verb is opted as the host for clitics despite the availability of other potential elements – marked by the underscore – for clitic hosting.

¹² In (38), following Sivandi's citic placement rule, direct or indirect objects which are marked by the differential object marking suffix $-\bar{a}$ are skipped for clitic hosting.

- (40) Semnani žo kola pevdā kard=**šon** do.pst=3pl:A 3SG.M.OBL:POSS hat visible 'They found his hat.' (PS. 26) (41) Lari **oš**=čī vekiveki imiva-yā bā degat one.by.one fruit-pl with care 3sg:a=pick.pst 'He pecked the fruit one by one with care.' (PS2.3) Nowdani (42) golābi-al а bālā-v deraxt eš=mi-či from pear-PL top-ez tree 3sg:a=IPFV-pick.pst 'He was pecking the pears on the tree.' (PS. 3) (43) Bastaki **š**=a-det āhangar ševal shovel blacksmith 3sg:r=IND-give.prs 'The blacksmith gives him a shovel.' (RS. 27) Likewise, proclitic attachment (enclitic attachment in the case of Semnani) is the
- (44) Nowdani om=ne-mi-šā 1sg:nc=neg-ind-can.prs 'I can not (come out).' (CG. 4) (45) Yazdi Zoroastrian in di hemla he-kr-ā be mi boz-ā, 3sg attack IRR-do.prs-3sg ADD to DEM goat-PL **šo**=be-xr-ā 3pl:0=IRR-eat.prs-3sg 'That he (too) attack these goats, (and) eat them.' (SM2. 6)

only option when a verb form is the sole element in the clause.

- (46) Lari **oš**=nā-yr-a 3sg:0=NEG-let.PRS-3sg 'He does not let it.' (PS1. 9)
- (47) Semnani ba-di=**šon** PFV-see.PST=3PL:A 'They saw.' (PS. 22)

Note that a vocalic element precedes the singular form of the person clitics in (41), (42), (44), and (46). This vocalic element is later argued to be the reanalysed form of the particle u- of Middle Iranian, which is now partly merged in the paradigm of person clitics in V-based clitic systems. In Section 6, we will argue that the movement of clitics toward the verb resulted in the reanalysis of the erstwhile S2-assuring particles and the actualization of the unit 'particle=clitic' as a proclitic in the now V-based clitic systems. Dabir-Moghaddam (2008) and Jügel (2017) share our view in considering this vocalic element as a reflex of the MWI particle u-. However, their analyses differ from the one outlined in this paper (see Section 5).

Finally, a few V-based proclitic systems illustrate some relics of erstwhile clausal second positioning. The difference is that the forms that are now pro-clitics appear clause-initially. The most straightforward examples of this are seen below. These examples further point to the mobility of the bound elements which we consider clitics in this paper.

(48)Bastaki **š**=a te kesa e-ke sack 3sg:a=into IPFV-do.PST 'He would put (the pears) into a sack.' (PS. 5) (49) Yazdi Zoroastrian **šo**=šuru kā pākre tamiz kārtā 3PL:A=start do.pst kitchen clean do.INF 'They started cleaning the kitchen.'

(BO. 12)

In the above examples, the A-past clitics are expected to occur on the light verb 'do'. However, they occur clause-initially, but in the guise of a proclitic. In Section 7 we argue that the proclitic attachment of person clitics in such contexts can be accounted for with



Figure 3: Procliticization and encliticization of pronominal clitics in WILs.

reference to the fact that historically particles would host person clitics clause-initially under certain conditions (i.e. in the absence of other clitic-hosting elements).

4 The extent of proclitic attachment across West Iranian

As mentioned earlier, a number of modern languages have developed proclitic attachment out of the erstwhile enclitic attachment of MWI. Figure 3 illustrates languages which have developed proclitic attachment.

Among the languages studied, 19 have enclitic attachment, while 12 languages have developed proclitic attachment along with enclitics. The map also shows that proclitic attachment is most conspicuously a feature of languages of Southeast Iran, Central Plateau languages, and Nowdani in the Southwest group. The languages investigated here which show proclitic attachment differ, however, with regard to the extent to which proclitics are used. In Table 2, the extent of proclitic attachment is examined on different hosts.

Language	Domain of cliticization		Host	
		Verb stem	там-verb stem	Preposition
Meymei	VP	+ (rare)	_	_
Badrudi	VP	+ (rare)	_	_
Khansari	VP	+ (rare)	+	_
Delijani	VP	_	+	_
Abuzeydabadi	VP	_	+	_
Naeini	VP	+	+	_
Yazdi Zoroastrian	Mainly V	+	+	+
Lari	Mainly V	+	+	+
Bastaki	Mainly V	+	+	+
Nowdani	V	+	+	+
Bandari	V	+	+	_
Minabi	V	+	-	+

 Table 2: Procliticization extent across investigated languages.¹³

As with pro-clitic attachment on bare verb stems, the languages are divided into three groups: the first group comprises all V-based clitic systems and Naeini, which has a VP-based clitic system. Here, proclitic attachment on the verb is the norm. Examples follow:

(50) Lari
bič-iā šo=got
child-PL 3PL:A=say.PST
'The children said.'
(SM. 6)

(51) Bastaki
 oš=di 3sG:A=See.PST
 'He saw.'
 (PS. 19)

¹³ Keys: +: proclitic attachment is allowed; -: proclitic attachment is not allowed.

- (52) Yazdi Zoroastrian kosapošt oš=vā turtle 3sg:A=say.pst
 'The turtle said.' (KX. 14)
- (53) Nowdani *eš=xard-e* 3sg:A-eat.PST-PRF 'He has eaten.' (SM. 34)
- (54) Naeini

ā **m**=*e*-*di*-*en* yes 1sg:A=TAM-see.PST-3PL:0 'Yes, I saw them.' (EL. 44)

The second group consists of Khansari, and (less so) Meymei and Badrudi. Here, proclitic attachment on bare verbs is only possible with certain verb forms, e.g. 'say', 'want'.

- (55) Khansari *ež*=*vāt* 3sg:A=say.pst 'He said.' (QB. 8)
- (56) Badrudi *ašun=vā* 3PL:A=**say**.PST 'They said.' (EL. 43)
- (57) Meymei *am*=gā hā-gir-on
 1sG:NC=want.PST PVB-IRR.take.PRS-1sG
 'I wanted to buy (it).'
 (EL. 69)

In the third group, clitics are not realized as proclitic on the bare verb. This concerns Delijani, Abuzeydabadi, and most verbs of the languages of the previous group.¹⁴

(58)	Delijani		
	ba= šūn -berd	ru_ve	garmua-i
	PFV=3PL:A-take.PST	inside	public.bath-INDF
	'They took him to a (GX. 33)	a public l	bath.'
(59)	Abuzeydabadi		

59) Abuzeydabadi
māsu=a ba=m-xard-a
fish=2sg:poss pFv=1sg:A-eat.pst-3sg.F
'I ate your fish.'
(BS. 16)

As with proclitic attachment on verb forms with a preceding TAM prefix, the languages are classified into two groups. The first group consists of the majority of V-based clitic systems, and VP-based clitic systems of Khansari, Delijani, and Abuzeydabadi. Here, the clitic attaches to the TAM prefix as a proclitic.

(60)	Delijani			
	āw	ašon =a-bar-a		
	water 'The wa	3PL:O=IND-take.prs-3	SG	
	(GX. 18)	ter win take them aw	uy.	
(61)	Khansar	i		
	šomā e	\check{z} =e-vin-di		
	2pl 3	SG:0=IND-See.PRS-2PL		
	'You see	him.'		
	(QB. 17)			
(62)	Bastaki			
	š =a-res-	et-e	peš-e	āsiābān
	3sg:0=ini	o-send.prs-3sg-drct	to-ez	miller
	'He send	ls him to the miller.'		
	(RS. 18)			

¹⁴ The diachronic development of TAM prefixes should be considered here. One hypothesis for the lack of proclitic attachment could be the grammaticalization of a preverbal particle as a punctual (perfective) prefix *ba-/be*- before the bare past verb stem across most of the Central Plateau languages (see MacKinnon 1977 for the development of *ba-*).

(63) Nowdani **mu**=mi-es 1PL:NC=IPFV-want.PST 'We wanted.' (EL. 69)

The procliticization tendency also holds if the preverbal element is a negative affix, cf. (64)–(66), or a preverbal derivational formative, cf. (67)–(69).

kot=eš

(64) Bastaki nun om=ne-xard-e bread 1sg:a=neg-eat.pst-prf 'I haven't eaten food.' (RS. 17) (65) Yazdi Zoroastrian *m*=*e*-*na*-*vā*t-*ā* 1sg:a=tam-neg-say.pst-prf 'I haven't said.' (EL. 9) (66) Nowdani *eš*=na-lešt 3sg:a=neg-let.pst 'He didn't let (the goat).' (PS. 9) (67) Nowdani bā sizan $\mathbf{\check{s}}=\bar{a}$ -doxt-a with needle 3sg:a=pvb-sew.pst-drct coat=3sg:poss 'He sewed (the list) with a needle to his coat.' (SL2. 21) (68) Bastaki *āšpazxune_pāk_* **om**=vā-kerd-e clean 1sg:a=pvb-do.pst-prf kitchen 'I cleaned the kitchen.'

(BO. 19)

(69) Lari **oš**=vā-düt 3sg:A=PVB-**sew**.PST 'She sewed (it).' (SM. 27)

Among V-based clitic systems, Bastaki, Lari, and Yazdi Zoroastrian are exceptions in that the non-verbal component of a complex predicate is treated like a TAM prefix and is procliticized upon.

- (70) Yazdi Zoroastrian $kosapošt \ um\bar{a} \ vo \ \check{s}e=qabul \ k\bar{a} \ /*qabul \ o\check{s}=k\bar{a}$ turtle come.PST and 3sG:A=acceptance do.PST 'The turtle came over and accepted (the challenge).' (KX. 10)¹⁵
- (71) Lari
 še=ejāza gete ke oču-a dar /*ejāza oš=gete
 3sg:a=permission take.pst compl go.prs.3sg-drct out
 'She asked for permission to go out.'
 (CG. 2)

In the above examples, the complex predicates have been procliticized upon. These examples could also be treated as cases of mobility of proclitics, and their freedom as regards host selection. Note that despite the apparent similarity to a VP-based clitic system, the clitic placement in ex. (70)–(71) is still best considered V-based positioning. Evidence in support of the V-based analysis comes from the following pair. In (72a) *salumalayk* 'greeting' is analysed as an object NP and is skipped as the clitic host (as opposed to a VP-based system). However, in (72b) it is analyzed as a light verb complement and is procliticized upon (cf. Mohammadirad 2020: Ch. 5 for the analysis of V-based and VP-based clitic systems).

(72) Yazdi Zoroastrian

a.	yaki	salumalayk	oš=kā	/*yaki	še=slumalayk	kā
	one	greeting	3sg:a=do.pst			

¹⁵ The Zoroastrian community in Yazd is assumed to have migrated to Yazd (Gholami 2016). The close parallels between procliticization in Yazdi Zoroastrian and Larestani dialects could possibly hint at the origin of the Zoroastrian community in Yazd in the South of Iran. Note further that proclitic attachment is a feature of the Zoroastrian dialect of Kerman as well (cf. Gholami 2016; Ivanow 1940).

b. še=salumalayk kā 3sg:a=greeting do.pst 'He said hello.' (HB1. 12)

As we shall explain in Section 7 these constructions can be reconstructed as residuals of the erstwhile clausal second positioning of enclitics.

The second group does not allow clitics to procliticize on the TAM prefix. This includes Minabi, Meymei, and Badrudi, cf. (73)–(75). Note that unlike the rest of the V-based clitic systems, Minabi does not allow proclitic attachment on the TAM prefix. This erratic behaviour is probably the result of contact influence from neighbouring Balochi, which has only enclitic attachment.

(73) Minabi mom-o bap=i a-go=šā /*šā=a-go mom-conj dad=3sg:poss IPFV-say.PST=3PL:A 'His parents would say.' (MM. 22)

- (74) Meymei *a=t-bard-on*IPFV=2sg:A-take.pst-1sg:o
 'You would take me.'
 (EL. 42)
- (75) Badrudi *to hem a=d-xor-on*2sg ADD IND=2sg:o-eat.PRS-1sg
 'I will eat you as well.'
 (SM1. 26)

Finally, as for cliticization on prepositions, the tendency for languages with the verb as the domain of cliticization is for clitics to procliticize on prepositions. Some examples follow:⁶

¹⁶ Note however that the extent of procliticization on prepositions is different among V-based clitic systems. It is only in Lari, Bastaki, and Yazdi Zoroastrian that proclitic attachment on prepositions is regularly at work. In other languages, e.g. Minabi, Nowdani, some prepositions have been borrowed from contact languages which have enclitic attachment. The borrowed prepositions are treated the same as in the source language for encliticization, hence the enclitic attachment on such prepositions.

- (76) Lari
 š=az_bar a_te sabad a-riz-en
 3sG:R=for in basket IND-pour.PRS-3PL
 'They put (the pears) into a basket for him.'
 (PS1. 18)
- (77) Yazdi Zoroastrian $d\bar{a}\check{s}t$ $\check{s}=e-k\bar{a}$ $\check{s}=e_tu$ hand 3sg:A-IPFV-do.PST 3sg:R=in'He put (his) hand in it.' (HB2. 12)
- (78) Minabi
 kār t=a hast=om
 job 2sg:r=to exist.prs=1sg:nc
 'I have a business with you.'
 (El. 70)
- (79) Nowdani *ye* bār dige t=aš mi-ga-m
 one time more 2sg:r=to IND-say.prs-1sg
 'I'm telling you again.'
 (EL. 21)

Proclitic attachment on prepositions is not an option in VP-based clitic systems or in Bandari, as exemplified below:

(80) Naeini dendeun na-dār-a ke ve=š hamla Ø-kir-a tooth NEG-have.prs-3sg COMPL to=3sg:r attack IRr-do.prs-3sg 'He has no teeth to attack her.' (SM. 39)

(81) Bandari
 zan=eš az=eš a-pors-ed woman=3sg:poss from=3sg:r IND-ask.prs-3sg
 'His wife asks him.'
 (SL2, 2)

As can be seen, the languages investigated here can be classified into different groupings with respect to the extent of proclitic attachment. What is evident,

though, is that while VP-based languages might have proclitic attachment in some restricted contexts, e.g. on verb forms with a preceding TAM, it is specially in the southern languages with V-based clitic systems that proclitics are the main apparatus for the phonological attachment of person clitics. Here, clitics in nearly all their different functions attach as proclitics. On the other hand, apart from Yazdi Zoroastrian, in Central Plateau languages proclitic attachment is limited to clitics which index core arguments, e.g. A and O: adpositional complement clitics and possessor clitics attach solely as enclitics, cf. Table 3. These differences between Central Plateau languages and V-based proclitic systems could indeed point to different historical trajectories in developing proclitics.

Clitic functions	en- or pro-clitic attachment	Languages
A-past and non-canonical subject indexing	Proclitic	V-based proclitic systems; certain verb forms in most Central Plateau
Object indexing	Proclitic	V-based proclitic systems; certain verb forms in most Central Plateau
	Enclitic	Badrudi, Meymei, Minabi, Nowdani
Adpositional complement	Proclitic	Bastaki, Lari, Yazdi Zoroastrian
indexing	Proclitic and enclitic	Minabi, and Nowdani
	Enclitic	Bandari, Central Plateau
Possessor indexing	Proclitic and enclitic Enclitic	Bastaki, Lari, Yazdi Zoroastrian Central Plateau, Minabi, Nowdani, Bandari

Table 3: The extent of proclitic attachment across different functions.

As seen in Table 3, among clitic functions subject indexing is the most liable to endorse proclitic attachment. This tendency becomes weaker for object indexing, and is only sporadically the case for adpositional complement clitics and possessor clitics. This further indicates that a hierarchy of proclitic attachment can be assumed for clitic functions, with subject indexing being the most likely to have proclitic attachment and possessor indexing being the least likely.

5 Previous scholarship on proclitic attachment in West Iranian

Proclitic attachment of clitics has been touched upon in passing in existing research (Dabir-Moghaddam 2008; Gholami 2018; Jügel 2017). However, no

thorough analysis, synchronic or diachronic, has, to our knowledge, so far been proposed. Consequently, we rely on glossing conventions used in these studies in order to grasp the gist of the underlying analysis. As previously mentioned, a crucial point to consider in the discussion of procliticization is the fact that some modern languages have developed the latter out of the previous enclitic attachment of clitics in the Middle Iranian period. In this transition, the particles *o*- and *a*-play an important role; both particles go back to clause-initial conjunctions *u*- and *a*(*h*)- in Middle Iranian and act as clitic hosts when other eligible clitic hosts are absent in the clause (see Sections 2 and 6). Dabir-Moghaddam (2008) and Jügel (2017) agree that the particle *o*- of modern languages in (82) originates in the Middle Iranian conjunction *u*. However, they treat *o*- differently:

(82) a. Lari/Davani

o=š got PTCL=3SG:A say.PST 'He said.' (Dabir-Moghaddam 2008: 94)

- b. Middle Iranian

 o=š vā
 PTCL=3SG:A SAY.PST
 'He said.'
 (Jügel 2017)

 c. Zoroastrian dialect of Kerman

 um=di
 - um=di 1sg:A=see.pst 'I saw.' (Gholami 2018: 117, transcription modified)

As the glossing of (82a) shows, Dabir-Moghaddam (2008) proposes the same treatment for the sequence *oš* as in the Middle Iranian period, where *o* is a clitic host. Synchronically speaking, this analysis works only for clause-based clitic systems (like Davani), but faces serious problems in V-based proclitic systems like Lari, cf. (83). Indeed, if *o* is still a clitic-hosting particle, it is expected that it will occur with all the cells in the clitic paradigm. This is not the case though: *o* only appears with the singular set of clitics. We will see in Section 6 that the retention of *o* before the singular set is motivated by syllabification restrictions.

(83)	Davani	i	Lari	
	o =m	di	o m=di	'I saw.'
	o =t	di	o t=di	'You saw.'
	0 =Š	di	o š=di	'S/he saw.'
	o =mu	di	mon=di	'We saw.'
	o =tu	di	ton=di	'You saw.'
	o =šu	di	šon=di	'They saw.'

Jügel (2017) takes the combination o + clitic as an oblique (or non-nominative) case form, cf. (82b) above. He considers the whole unit as an independent oblique pronoun.¹⁷ His view seems to be only applicable to Middle Iranian data. However, Jügel seems to take the same stance for the analysis of such a unit in Yazdi Zoroastrian, which has the same paradigm as the one from Lari seen in (83). Thus, the combination o + clitic can only be assumed for singular forms.

Finally, Gholami's (2018) glossing suggests that *u*- has been merged with the clitic pronoun. The analysis advocated for in this paper is in line with this view. The author attributes the rise of procliticization in modern languages to the loss of the ergative construction. This analysis is refuted here: there is no direct correlation between ergative decay and the rise of proclitics. For example, both Davani and Dashti have undergone ergative decay, yet neither has developed proclitics. In addition, the author assumes that only A-past clitics have become proclitics (Gholami 2018: 177). However, proclitic attachment involves virtually all clitic functions, cf. Table 3 above.

6 The fate of the clitic-hosting particles in modern languages

In this section we discuss the fate of clitic-hosting particles in languages which have developed proclitics. The underlying assumption is that historically, what are now proclitics were originally enclitics on some clitic-hosting particles (or conjunctions). Following the rightward drift of clitics (and/or their attachment to the head), the unit 'particle=clitic' was reanalysed as a proclitic on some host to the right. This development tightly correlates with the loss of the discourse-structuring function of conjunctions. The reanalysis of the unit 'particle=clitic' presumably happened after the rightward drift of clitics in the clause, and/or their attraction to the relevant heads. In what follows, the rise of proclitics will be investigated in the light of changes that occurred to the paradigm of enclitics.

We saw that the S2 cliticization system of WMI has survived to a large extent in a few modern WILs, such as Dashti and Davani. In both these languages, the reflex of the particle *u*- guarantees the second positioning of clitics, in case there is no appropriate host (cf. Section 3.1 for examples). In this section, we track the fate of the clitic-hosting particles in languages which developed proclitic attachment. In

¹⁷ Jügel's view seems to be inspired by Ivanow (1940: 64), who takes the clitic forms of the Zoroastrian dialects of Kerman and Yazd (same paradigm as Lari) as "independent personal pronouns". His stance is refuted here since the clitic forms are still prosodically deficient and cannot stand by themselves, e.g. in response to a question.

particular, we focus on the paradigm of clitics in these languages and the ways in which the particles have been integrated into the clitic paradigm as a result of the reanalysis. We argue that the reanalysis of clitic-hosting particles at some point in the history of what are now proclitic systems is a necessary condition for the rise of proclitics, as opposed to languages which went with enclitic attachment.

Reanalysis is one of the main mechanisms of syntactic change and is defined as follows (Langacker 1977: 58): "[A] change in the structure of an expression or class of expressions that does not involve any immediate or intrinsic modification of its surface manifestation". Reanalysis can affect different layers of structure, including constituency, hierarchical structure, category labels, grammatical relations, etc. (cf. Harris and Campbell 1995: Ch. 4).

Table 4 illustrates the development of S2-assuring particles in WILs. Notice that the particle *u*- can phonologically change into either *o*-, as in Davani, Dashti, or *e*-, as in the 2SG and 3SG forms of clitics in Dashti and Nowdani.¹⁸ Particle \bar{a} -/*a*- changes to either *a*- or *e*- through vowel raising, as in Meymei, Khansari, etc.

According to Table 4, only Davani and Dashti have fully preserved a reflex of uand/or a- particles in all persons. Here, the particles still guarantee the S2 requirement for enclitic positioning (see Section 3.1 for examples). On the other hand, in Khansari, Badrudi, Meymei, and Delijani, the erstwhile particle a- is now merged into the clitic paradigm in all cells (cf. [85] below). However, in the rest of the languages in Table 4, the erstwhile particles, now functioning as supporting vowels, are only retained in the singular set of clitics, hence *om*, *ot*, *oš* versus the plural set *mo*, *to*, *šo*. Note that the /o/ or /u/ vowels in the plural forms are indeed plural markers and should not be conflated with the supporting *o* which precede the person clitics.

Assuming the initial S2 clitic placement rule for the VP-based and V-based clitic systems, the question arises as to what kind of changes the particles have gone through before becoming part of the paradigm of clitics. We propose that the proximity of the sequence 'particle=clitic' to the verb stem finally led to a reanalysis of the *u*- and/or *a*- particles as part of the paradigm of person clitics. This shift appears to have happened posterior to the rightward drift of person clitics since the Middle Iranian period, and in general can be divided into three stages.

In the first stage, following the clause-second restriction on the placement of clitics, S2-assuring particles occur before all persons. Davani and Dashti have remained at this stage. This is illustrated below for the paradigmatic form of the verb 'to ask' in the simple past tense:

¹⁸ Indeed, it is possible that the vowel *e* in the V-based proclitic systems might have been derived from the particle *a*-.

Cliticization domain		Clause		Verb phrase				Verb		
Languages	IWM	Davani	Dashti	Badrudi/Meymei/Delijani	Khansari	Lari	Nowdani	Yazdi Zoroastrian	Bandari	Minabi
1s6	m=n	т=0	0=m	am=	em=	om=	om=	om=	om=	om=
2sg	u=t	o=t	e/o=t	at=	ed=	ot=	et=	od=	et=	et=
3sg	u=Š	0=Š	e∕o=š	aš=	ež=	oš=	eš=	oš=	eš=/ī=	Ē
1PL	u=mān	o=mū	o=mū	am un=	emun=	=om	mū=	<i>mo=</i>	=0W	=uom
2PL	u=tān	o=tū	e/o=tū	adun=	edun=	to =	tū=	do=	to =	ton=
3pl	u=šān	o=šū	e∕o=šū	ašun=	ežun=	š0=	šū=	šo=	š0=	šon=

Table 4: Reflexes of *u*- and *a*- in the clitic paradigms of modern Iranian languages.

(84)	Davani			
	o =m	porsi	[PTCL=1SG:A ask.PST]	'I asked.'
	o =t	porsi	[PTCL=2SG:A ask.PST]	'You asked.'
	0 =Š	porsi	[PTCL=3SG:A ask.PST]	'S/he asked.'
	0 =mu	porsi	[PTCL=1PL:A ask.PST]	'We asked.'
	o =tu	porsi	[PTCL=2PL:A ask.PST]	'You asked'
	0 =šu	porsi	[PTCL=3PL:A ask.PST]	'They asked.'

At stage 2, through boundary shift the bimorphic unit 'particle=clitic' (e.g. o=m *xward* 'I ate') is reanalysed as a unimorphic unit, hence *om*, *ot*, etc. In the absence of leftward support, this unimorphic unit is subject to actualization as a proclitic on the next element to the right, in this case the verb, hence, om=xward 'I ate'. Consequently, the enclitic attachment changes into a proclitic attachment. This shift is represented fully in the paradigm of person clitics in some Central Plateau languages, e.g. Khansari, Badrudi, Meymei, and Delijani.

(85) Badrudi $am = v\bar{a}$ [1sg:A=say.pst]

a m=vā	[1sg:a=say.pst]	'I said.'
a d=vā	[2sg:A=say.pst]	'You said.'
a š=vā	[3sg:a=say.pst]	'S/he said.'
a mun=vā	[1PL:A=say.PST]	'We said.'
a dun=vā	[2PL:A=say.PST]	'You said.'
a šun=vā	[3pl:a=say.pst]	'They said.'

Southern languages have also gone through stage 2, i.e. a relic of a clitic-hosting particle is retained in the clitic paradigm. Data from Buringuni, a south-western language of Fars province, call for a partial retention of erstwhile clitic-hosting particles before all person values in the clitic paradigm (*om*, *et*, *eš*, $[o]m\bar{u}$, $[e]t\bar{u}$, [e] $s\bar{u}$). Buringuni has a V-based clitic system. In (86)–(87) it can be seen that a reflex of the clitic-hosting particle occurs with the 1SG and 1PL clitic forms.

(86) Buringuni

kār-iom=hanbusiness-INDF1sg=exist.prs'I have a business (to do).'(Mann 1909: 99, transcription modified)

(87) Buringuni

māhamomu=zay1PLtoo1PL:A=hit.PST'We shot too.'(Mann 1909: 91, transcription modified)

Finally, stage 3 highlights the shift in the conditioning factor for the occurrence of the erstwhile particle, namely, ensuring that the process of cliticization would not violate the syllable structure rules of the language. The vocalic *o* is retained in the singular set to avoid non-licensed onsets (**mx-*, *tx-*, *šx*). The syllabic plural clitics do not need to resyllabify with *o*.

(88) Lari

o m=xa	/*mxa	[1sg:a=eat.pst]	'I ate.'
o t=xa	/*txa	[2sg:A=eat.pst]	'You ate.'
o š=xa	/*šxa	[3sg:A=eat.pst]	'S/he ate.'
mo=xa		[1PL:A=eat.PST]	'We ate.'
to=xa		[2PL:A=eat.PST]	'You ate.'
šo=xa		[3pl:A=eat.pst]	'They ate.'

The data thus suggest that, as a result of reanalysis, the erstwhile S2-assuring particles have gradually fused into the paradigm of person clitics, and later disappeared from the plural set.¹⁹ These changes are assumed to have occurred following the rightward drift of clitics towards the verb, and are summarized in Table 5.

	1st stage	2nd stage	3rd stage
1sg	e/o =m	e/o m=	e/o m=
2sg	e/o= t	e/ot=	e/o t=
3sg	e/o= š	<i>e/o</i> š=	e/o š=
1pl	e/o= mu	e/omu=	mu=
2pl	e∕o= tu	e/o tu=	tu=
3pl	e/o= šu	e/o šu=	šu=

Table 5: Presumed stages of the development of the *u*- and *e*- particles before the bare verbstem.

19 Despite the general lack of reflexes of the *o*- particle before plural clitics in V-based clitic systems, *o* can nevertheless sometimes reappear with plural forms in these languages, especially when two identical or similar person clitics follow each other, e.g. $hav\bar{a} \cdot y xo = to otu = bu$ [weather-EZ self=2PL:POSS 2PL:NC=be.IMP] 'Take care of yourselves.' [lit. 'hold your weather'] (Nowdani. SM. 3). Here, the existence of the supporting vowel *o* before the 2PL clitic shows that the *o*-vowel can recur with plural clitics as well, where it was historically present. Alternatively, the occurrence of the vowel *o* before the plural clitic can be motivated by the strategy of 'avoidance', a tool used by the grammar to preclude the repetition of identical morphemes in a row (see Menn and MacWhinney 1984; Yip 1998). In any case, this example suggests that the underlying supporting *o* can be postulated in the onset of the plural set of clitics as well, but in the current state of Nowdani it only appears under certain morphophonological conditions.

One can retell the dynamics of this shift as follows: the type of reanalysis is the loss of a morpheme boundary, so that the bimorphic unit 'particle=clitic' in o=m, o=t, o=š was reanalysed as a single morpheme: om, ot, oš. These single morphemes are then subject to actualization: they manifest a new syntactic behaviour by appearing on the verb as a proclitic. The cause of this change is assumed to be the abandoning of clause-second positioning as the domain of cliticization in the majority of languages. The unit 'particle=clitic' was thus set free from the S2 requirement to be placed clause-initially. Consequently, it approached the verb and gradually developed into a single morpheme, yet in the absence of leftward support it was actualized as a proclitic on the verb. Finally, the effect of this reanalysis was: first, the loss of the discourse-structuring function of conjunctions and their bleaching as S2-assuring particles in Middle Iranian; second, the change in contexts where the S2-assuring particle would appear: from clause-initial position (in Middle Iranian, Dashti, and Davani) to the verbal domain (e.g. in Lari, Nowdani, Yazdi Zoroastrian).

We could go further and assume that the S2-assuring particles have undergone the same development before TAM forms of verbs. In the presumed earlier stage, *u*- and *a*- particles would host the clitics clause-initially when the verbal form was the only available element in the clause. This is seen below (see also ex. [32] above):

Dashti **o=mu** mi-košt-an PTCL=1PL:A IPFV-kill.PST-3PL:0 'We would kill them.' (EJ. 20)

(89)

A paradigmatic form of the verb 'can, be able' in the present tense of Davani further exemplifies the first stage of development:

(90)	Davani	Davani							
	o =m	mi-šā	[PTCL=1SG:NC TAM-be able.prs]	'I am able.'					
	o =t	mi-šā	[PTCL=2SG:NC TAM-be able.prs]	'You are able.'					
	0 =Š	mi-šā	[PTCL=3sg:nc tam-be able.prs]	'S/he is able.'					
	0 =mu	mi-šā	[PTCL=1PL:NC TAM-be able.prs]	'We are able.'					
	o =tu	mi-šā	[PTCL=2PL:NC TAM-be able.prs]	'You are able.'					
	0 =šu	mi-šā	[PTCL=3PL:NC TAM-be able.prs]	'They are able.'					

In the next stage, these particles are reanalysed as part of the clitic paradigm, resulting in the univerbation of the single unit. The data from the VP-based clitic system of Delijani represents this stage:

(91)	Delijani		
	a m=e-gā	[1sg:nc=ipfv-want.pst]	'I would wish.'
	a t=e-gā	[2sg:nc=ipfv-want.pst]	'You (sg.) would wish.'
	a š=e-gā	[3sg:nc=ipfv-want.pst]	'S/he would wish.'
	a mon=e-gā	[1pl:nc=ipfv-want.pst]	'We would wish.'
	a ton-e-gā	[2pl:nc=ipfv-want.pst]	'You (pl.) would wish.'
	a šon=e-gā	[3PL:NC=IPFV-want.PST]	'They would wish.'

In the same way, southwestern Buringuni exemplifies stage 2, i.e. the retention of erstwhile particles before all person values:

(92)	Buring	Buringuni						
	kujā	šomā	e tū=mi-ā	huš-īt?				
	where	2pl	2PL:NC=IND-want	go.prs.irr-2pl				
	'Where	'Where do you want to go?'						
	(Mann	(Mann 1909: 96, transcription modified)						

In the last stage of development, the recourse to the supporting vowel *o* depends on the form of the TAM prefix. If the latter is vocalic, the singular clitics can resyllabify with it. Consequently, the supporting vowels disappear from the paradigm of clitics.

(93) Lari [1SG:A=IPFV-eat.PST] 'I was eating.' m=a-xa[2sg:A=IPFV-eat.Pst] 'You were eating.' t=a-xa 'S/he was eating.' š=a-xa [3sg:A=IPFV-eat.Pst] [1PL:A=IPFV-eat.PST] 'We were eating.' mo=a-xa[2PL:A=IPFV-eat.PST] 'You were eating.' to=a-xa šo=a-xa [3PL:A=IPFV-eat.PST] 'They were eating.'

However, if the TAM marker is not vocalic, but a consonant-initial syllabic unit, *o* is retained before the singular set. Therefore, the process of cliticization complies with the syllable structure rules of the language:

(94)	Nowdani			
	o m=mi-xa	*mmi	[1sg:a=ipfv-eat.pst]	'I was eating.'
	e t=mi-xa	*tmi	[2sg:A=IPFV-eat.Pst]	'You were eating.'
	e š=mi-xa	*šmi	[3sg:a=ipfv-eat.pst]	'S/he was eating.'
	mu=mi-xa		[1pl:A=IPFV-eat.pst]	'We were eating.'
	tu=mi-xa		[2PL:A=IPFV-eat.PST]	'You were eating.'
	šū=mi-xa		[3PL:A=IPFV-eat.PST]	'They were eating.'

Likewise, proclitic attachment to the TAM form of the verb in some VP-based clitic systems (e.g. Abuzeydabadi, Naeini) is representative of this stage:

(95)	Abuzeydabadi						
	m=a-xand	[1sg:a=tam-read.pst]	'I was reading.'				
	d=a-xand	[2sg:a=tam-read.pst]	'You were reading.'				
	y=a-xand	[3sg:a=tam-read.pst]	'S/he was reading.'				
	mon=a-xand	[1pl:a=tam-read.pst]	'We were reading.'				
	don=a-xand	[2PL:A=TAM-read.PST]	'You were reading.'				
	yon=a-xand	[3pl:a=tam-read.pst]	'They were reading.'				

The stages of the development of the *u*- and *a*- particles before TAM forms of verbs are summarized in Table 6.

The data thus highlight a gradual attachment of the erstwhile particles to the paradigm of clitics. This shift results from the reanalysis of the unit 'particle=clitic' and the ensuing merging of the particle into the clitic paradigm, which paved the way for the actualization and the univerbation of the stray clitic in the form of a proclitic.

To sum up, the specific claim we make here is that the rise of proclitics on verbal forms in languages with proclitic attachment in Table 4 is directly related to the fact that the reflexes of the *u*- and/or *a*- particles were reanalysed as a part of the paradigm of person clitics. As sketched above, this change is a gradual process and is assumed to have probably occurred after the rightward drift of person clitics towards the verb. The fact that the verb shows certain attraction for grammatical meaning (Steele 1977) further facilitated the reanalysis. Consequently, the conditioning factor for the retention of such particles (that is, to guarantee that clitics have S2 positioning) was no longer valid, since clausal second positioning of person clitics had already given way to VP-based and V-based clitic systems. The old particles are now fully or partly integrated into the paradigm of person clitics:

	1st stage	2nd stage	3rd stage
1sg	e/o =m	e/o m=	m= / o m=
2 sg	e/o=t	e/o t=	t= / o t=
3 sg	e/o= š	e/o š=	š= / o š=
1pl	e/o =mu	e/omu=	mu=
2pl	e∕o =tu	e/o tu=	ut=
3pl	e/o =šu	e/o šu=	šu=

Table 6: Presumed stages of the development of the *u*- and *a/e*- particles before TAM forms of verbs.

having occurred with all the person values in the earlier stages, they were dropped from the plural forms at a later stage, to remain only on singular forms – solely to assure that the process of cliticization would not violate the syllable structure rules of the languages.

What we observed so far was the rise of proclitics in the immediate preverbal domain. The earlier enclitic stage is also assumed for other clitic functions and for the positioning of clitics in the non-preverbal domain. In the following examples, which represent an earlier state of affairs, the particle hosts a fronted adpositional complement clitic, and a possessor clitic (or an indirect affectee), respectively.

(96) Davani

 $o=\check{s}$ jaryān $a\check{s}$ mi-ga-tā PTCL=3SG:R Story to IND-tell.PRS-3SG 'He says the story to him.' (KS. 21)

(97) Middle Persian

u=tazpus täbräd_wistudsePTCL=2SG:POSSfrom son till brother twenty and threemurdbawēnddeadbe.PRS.3PL'And of your sons up to your brothers twenty-three will be dead.'(Durkin-Meisterenst 2014: 327, mpB. 400)

These examples not only show that the adpositional complement clitics and the possessor clitics abide by second positioning, but also indicate that the clitic-hosting particle *o*- is a relevant host for such clitic functions. The following examples further illustrate the occurrence of the unit 'particle=clitic' in contexts not immediately preceding the (light) verb.

(98) Davani

o=**t** ya memuni hā-de PTCL=2sG:R One party PVB-give.PRS.1sG 'That I throw a party for you.' [lit. 'that I give you a party'] (XX. 14)

(99) Middle Persian

u=šomēragguftPTCL=3sG:Atoyoung.mansay.Pst'He said to the young man.'(Durkin-Meisterenst 2014: 275, mpB. 60)

(100) Davani **o**=š bad me-am-a i baček-e PTCL=3SG:NC bad IPFV-come.PST.3SG-DRCT DEM child-DEM1 'She hated this kid.' (KS. 8)

In short, (96)–(100) display the earlier state of affairs in which the particle element would host the second position clitics in their different functions, also including adpositional complement clitics and possessor clitics. In what follows, in what are now proclitic systems, the shift from second position clitics to proclitics is observed for adpositional complement clitics and possessor clitics. In other words, these clitic functions were subject to head attraction and ended up attaching to their heads (cf. Section 7).

(101) Nowdani

t=ašmi-ga-m2sg:R=toIND-say.PRS-1sg'I tell you.'(EL. 21)

(102) Lari

m=a_vāzyādčed-e1sg:poss/Nc=frommemorygo.pst-prf'I have forgotten.' [lit. 'it has gone from my memory'](EL. 56)

The following examples further show that the erstwhile 'particle=clitic' is now actualized as a proclitic clause-initially.

(103) Bastaki
šon=a_te sabad nā
3PL:A=in basket put.PST
'They put (the pears) into the basket.'
(PS. 15)

(104) Yazdi Zoroastrian *kosapošt umā vo še=qabul kā /*qabul oš=kā*turtle come.PST and 3sG:A=acceptance do.PST
'The turtle came over and accepted (the challenge).'
(KX. 10)

In the next section we elaborate on these issues and suggest that the current proclitics in their different functions and in different contexts can be ultimately analysed as deriving from the clausal second positioning of enclitics.

7 Procliticization as a residual of second positioning

The previous sections discussed in detail the process of integration of clitic-hosting particles into the paradigm of clitics as a result of reanalysis and actualization. It was held that through boundary shift the unit 'particle=clitic' became a single unit, and actualized as a proclitic on some host to the right. This section highlights the rise of procliticization in (mostly) non-verbal domains by making a parallel to the clitic systems in which particles hold clitics in parallel constructions. We will argue that the erstwhile presence of the clitic-hosting particles in the clause-initial position can account for the now clause-initial proclitics.²⁰

First, consider that neither in clause-based clitic systems (which have clitichosting particles), nor in V-based proclitic systems, are pre-verbal TAM formatives interrupted for clitic hosting:

(105)	Dashti	Dashti					
	o=mu	mi-košt-ar	1				
	PTCL=1PL:A	IPFV-kill.PS	t-3pl:0				
	'We would	kill them.'					
	(EJ. 20)						
(106)	Nowdani						
	mu =mi-es	<*o=mu	mi-es				
	1pl:nc=ipfv-	want.pst					
	'We wante	d.'					
	(EL. 69)						

Building on such a parallel, one may further assume that the proclitic on the TAM prefix previously had the particle as its host, (an option still available for Dashti in [105]). However, with the verb attracting the clitic in (106), the particle lost its clitic-hosting function and was lost clause-initially. The stray clitic then had to phonologically attach to the verb in a proclitic grab.

²⁰ A rather different path for the rise of proclitics is seen in Ossetic, an Eastern Iranian language. The latter has developed possessive proclitics under contact influence from West Caucasian languages (cf. Erschler 2009).

Another piece of evidence supporting the derivation of proclitics from the previous enclitic stage comes from cliticization on complex predicates. Recall from Section 4 that this is a feature of Yazdi Zoroastrian, Lari, and Bastaki. In the clause-based systems, the particle appears before the light verb complement to guarantee the second positioning of the clitic, cf. (107)–(108). In the verb-based system in (109), on the other hand, the clitic procliticizes to the light verb complement. Here again, an earlier stage is presumed in which the particle would hold the erstwhile enclitic in the clause-second position.

(107)	Middle Pers	Middle Persian							
	u=t [dašn d	ād] ^{CP}						
	PTCL=2SG:A g	gift g	ive.pst						
	'You gave g	ift.'							
	(Durkin-Mei	isterens	t 2014: 426	, mpB.928)					
(108)	Davani								
	0 =Š	[bad	me-am] ^{CI}	°-a	i	baček-e			
	ptcl=3sg:nc	bad	IPFV-come	e.pst.3sg-drct	DEM	child-дем1			
	'She hated t	this kid	.'						
	(KS. 8)								
(109)	Bastaki								
	ma =[xaš	$ez\bar{a}]^{Cl}$	P21	<*o=m x	caš e	zā			
	1sg:nc=nice	IND.CC	me.prs.3sg						
	'I like (to pl	ay with	my fish).'	[lit. 'my plea	sure co	mes']			
	(BS. 3)								

A third context for the derivation of proclitics from erstwhile enclitics is the attachment of a possessor clitic to the prepositional head of a PP which governs the possessor NP, cf. (110). By contrast, (111)–(112) illustrate cases of S2-positioning of the possessor clitic (or an indirect affectee), where the latter is attached to the particle *u*- in the sentence-initial position. Indeed, this parallel suggests that while the proclitic is frozen on the preposition in (110), it would appear clause-initially on the particle in an earlier stage, cf. (111)–(112).

²¹ However, assuming an earlier S2-stage still leaves open the challenge brought about by the presence of the vocalic a on the 1SG clitic, hence ma=. The vocalic a- could be considered a secondary development here, i.e. probably a sort of particle existed before the complex verb, which later disappeared.

(110) Bastaki mehr-e š=a te del a-kat dot-u. affection-EZ girl-def 3sg:poss=in heart IND-fall.prs.3sg <*mehre dot-u, o=š a te del kat 'He was filled with the affection for the girl.' [lit. 'the affection of the girl, fell into his heart'] (PD. 26) (111) Middle Persian

u = ttä bräd wist az pus ud se PTCL=2SG:POSS from son till brother twentv and three murd bawēnd dead be.prs.3pl 'And of your sons up to your brothers twenty-three will be dead.' (Durkin-Meisterenst 2014: 327, mpB. 400)

(112) Davani

 o=m
 az
 yād_
 še-s-e

 PTCL=1SG:POSS
 from
 memory
 g0.PST-EP-PRF

 'I have forgotten.' [lit. 'it has gone from my memory']

 (EL. 56)

Another context for the assumed derivation of proclitics from earlier clause-second enclitics is the proclitic attachment of a prepositional complement clitic on its head:

(113) Lari
š=az_bar a_te sabad a-riz-en <*o=š az_bar a_te sabad arizen
3sg:r=for in basket IND-pour.PRS-3PL
'They put (the pears) into a basket for him.'
(PS1. 18)

(114) Nowdani t=aš mi-ga-m <*o=t aš mi-ga-m 2sg:R=to IND-say.PRS-1sg 'I will tell you.' (EL. 21)

The examples just seen are comparable to the parallel constructions in (115)–(116) below. Here, the particles host the otherwise hostless prepositional complement

clitics. Based on this parallel, in (113)–(114) above the clitics were attracted to their heads, and lost their earlier hosts, i.e. the particles.

(115)	Middle Persian									
	u=t	dard	ud	danāh	abar_	nē	rasēd			
	PTCL=2SG:R	pain	and	illness	upon	NEG	arrive.prs.3sg			
	'And pain a	'And pain and illness does not come over you.'								
	(Durkin-Me	eisterens	t 2014	: 303, mp	oT. 220)					
(116)	Davani									
	0 =Š	jaryān	aš_	mi-ga-	tā					
	PTCL=3SG:R	story	to	IND-tel	1. prs-3sg					
	'He says th	'He says the story to him.'								
	(KS. 21)									

The same analysis could be applied to the procliticization of clitics on the preposition head of a prepositional phrase which precedes the verb. This was seen to be a feature of Lari and Bastaki. Contrast the following parallel constructions from Middle Persian and Bastaki:

(117)	Middle Pe	rsian						
	u= š	0	mērag		guft			
	PTCL=3SG:A	to	to young.n		say.pst			
	'He said to the young man.'							
	(Durkin-Meisterenst 2014: 275, mpB. 60)							
(118)	Bastaki							
	$\check{s}=a$	mamā	i=š	got	<*o=š a mama=š go	t		
	3sg:a=to	mom=	-3sg:poss	say.	PST			
	'She said	to her 1	nom.'					
	(CG. 13)							

To sum up, both S2-based clitic systems and V-based proclitic systems were argued to have preserved a reflex of the MWI particle *u*-. The modern particle *o*-, while having preserved its older function in a few modern languages, is now only traceable in the singular set of clitic paradigms of V-based proclitic systems (when clitics attach to the verb), but has totally disappeared before clitics clause-initially. By assuming an older stage in which the particle *o*- was present in the clause-initial position of V-based clitic systems, and through comparison with clause-based clitic systems, we arrived at the conclusion that proclitic attachment in the non-verbal domain could be derived from (i) the loss of the particle *o*- in such a domain, and (ii) the actualization of the stray clitics as proclitics on some hosts to the right.

Keeping the larger picture of cliticization in WILs in mind, the question arises as to why the majority of Iranian languages have stuck to enclitic attachment despite abandoning the clausal second positioning of person clitics. The general preference for enclitics over proclitics (Cysouw 2005) seems a possible explanation, but it provides little information on why enclitic attachment survived in the majority of languages. An alternative scenario would be diachrony and the branching of languages. That is, languages which preserved enclitic attachment have probably grammaticalized a more syntactic version of second positioning, a version in which the role of clitic-hosting particles was trivial as anchoring elements. For this purpose, investigation into Middle Iranian data can illuminate the derivation of languages which preserved enclitic attachment. For instance, Brunner (1977: 108) holds that the particle *u*- is used less frequently in Parthian than in Middle Persian. The languages with enclitic attachment then might possibly have descended from a Middle Iranian language, e.g. Parthian, in which the role of the clitic-hosting particle u- as a clitic host was insignificant or irrelevant.

8 Conclusion

In this paper, we surveyed the extent of the proclitic attachment of person clitics in West Iranian languages and outlined some hypotheses for its rise. We argued that the proclitics in the current VP-based and V-based clitic systems could be derived from an earlier stage in which clitics occurred in the clause-second position as enclitics. In other words, what are now proclitic systems are a residual of the erstwhile clausal second positioning of enclitics. It was held that proclitic attachment occurred after the rightward movement of clitics from clause-second position and their attachment to relevant heads. Through this drift, the S2-assuring particles were either lost or reanalysed as part of the paradigm of clitics, leaving clitics bereft of left-hand support for their realization. The stray clitics then actualized as proclitics on some host to their right.

Put differently, the proclitic attachment of clitics in WILs can be seen as an effort on the part of the clitic system to approach the head, e.g. the verb or preposition. The same analysis is suggested for proclitic attachment in some Uto-Aztecan languages in Steele (1977), where it is argued that proclitic attachment on the verb is a secondary development from second positioning with enclitic attachment. What we observe here is also reminiscent of the rise of proclitics in Old Romance; as put by Wanner (1987: 237): "Proclisis results from a lack of a lefthand prosodic support for the second position weak element or pronoun". Therefore, it might be safe to say that the rise of proclicis in WILs is another illustration of the

directionality of change in clitic systems, with S2 enclitics ending up as verbal affixes realized as proclitics.

Abbreviations in glosses²²

past transitive subject
additive
comparative
complementizer
conjunction
copula
complex predicate
definite
demonstrative
demonstrative particle
differential object (or indirect object) marking
directional particle
epenthesis
ezafe
feminine
imperative
indicative
indefinite
infinitive
imperfective
irrealis
light verb complement
masculine
non-canonical subject
negative
object
oblique
perfective
plural
possessor
perfect
present
past
particle
participle
preverbal derivational formative
adpositional complement or non-flagged secondary object

²² Glosses follow The Leipzig Glossing Rules (https://www.eva.mpg.de/lingua/pdf/Glossing-Rules.pdf), to which we added some language-specific labels.

REL relativizer s2 second position TAM tense-aspect-mood affix

Acknowledgment: We thank Geoffrey Haig for his comments on the draft version of the article. We also express our gratitude to the two anonymous reviewers for their comments, which led to significant improvement of the paper. We are grateful to Emmanuel Giraudet from the research group 'Centre de recherche sur le Monde Iranien' for the creation of maps. Any remaining shortcomings remains our own. Masoud Mohammadirad's work was enabled by the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement No. 665850.

Appendix

In Table 7 some information is provided about each language investigated in this paper. Due to limitations of space only the number of spoken narratives have been provided in the dataset column (cf. Mohammadirad 2020: Ch. 1 for metadata). Note that as there is no reliable census on language communities it is hard to provide realistic data for the number of speakers, given also the fact that a considerable portion of many language communities grow with little knowledge of native languages. Therefore, for some languages the number of speakers can alternatively be interpreted as the population size.

Size of dataset	Approximate n. of speakers	Location	Further information
4 spoken	100,000	35.99534,	cf. Sorani (<central kurdish)<="" td=""></central>
narratives		45.88214	in Glottolog
3 spoken	800,000	35.31271,	cf. Sine'i (<central kurdish)<="" td=""></central>
narratives		46.99745	in Glottolog
4 spoken	30,000	35.87820,	cf. Garruci (<southern< td=""></southern<>
narratives		47.59137	Kurdish) in Glottolog
4 spoken	5,000	35.25539,	cf. Hawraman-i Taxt (<gur-< td=""></gur-<>
narratives		46.25510	ani) in Glottolog
3 spoken	2000	35.13358,	Not listed in Glottolog
narratives		47.80199	
4 spoken	100,000	34.26456,	cf. Laki in Glottolog
narratives		47.61051	-
	Size of dataset 4 spoken narratives 3 spoken narratives 4 spoken narratives 3 spoken narratives 4 spoken narratives 4 spoken narratives	Size of dataset of speakers of speakers 4 spoken 100,000 narratives 3 3 spoken 800,000 narratives 4 4 spoken 30,000 narratives 4 4 spoken 5,000 narratives 3 3 spoken 2000 narratives 4 4 spoken 100,000 narratives 4 4 spoken 100,000 narratives 4 4 spoken 100,000	Size of dataset Approximate n. of speakers Location 4 spoken narratives 100,000 35.99534, 4 spoken arratives 100,000 35.99534, 3 spoken 800,000 35.31271, narratives 46.99745 46.99745 4 spoken 30,000 35.87820, narratives 47.59137 4 spoken 5,000 35.25539, narratives 46.25510 3 spoken 2000 35.13358, narratives 47.80199 4 spoken 100,000 34.26456, narratives 47.61051

 Table 7: General information on investigated languages.

Language	Size of dataset	Approximate n. of speakers	Location	Further information
Laki Harsini	7 spoken	20,000	34.27218,	cf. Belelli (2016); Lakic
	narratives		47.58576	Southern Kurdish in
				Glottolog
Chali (Shali)	3 spoken	15,000	35.89884,	cf. Southern Tatic (<tatic) in<="" td=""></tatic)>
	narratives		49.76503	Glottolog
Takestani	1 spoken	50,000	36.07556,	cf. Takestani (<ramand-< td=""></ramand-<>
	narrative		49.69086	Karaj < Tatic) in Glottolog
Semnani	2 spoken	100,000	35.59008,	cf. Semnani-Biyabuneki in
	narratives		53.37433	Glottolog
Central Taleshi	1 spoken	200,000	37.79848,	cf. Central Talyshi in
	narrative		48.90740	Glottolog
Delijani	2 spoken	20.000	33.99254.	cf. Delijani (<soic) in<="" td=""></soic)>
	narratives		50.68019	Glottolog
Khansari	2 spoken	15,000	33.29052,	cf. Khunsaric in Glottolog
	narratives		50.31922	C
Meymei	3 spoken	6.000	33.44972.	cf. Mavma'i (<soic) in<="" td=""></soic)>
	narratives		51.16804	Glottolog
Abuzevdabadi	2 spoken	6.000	33.89916.	cf. Abuzevdabadi (<soic) in<="" td=""></soic)>
,	narratives		51.76530	Glottolog
Badrudi	4 spoken	14.000	33.69181.	cf. Badrudi (<natanzic) in<="" td=""></natanzic)>
	narratives	.,	52.00395	Glottolog
Nikabad-Iondun	5 spoken	6.000	32.30685.	cf. Gazic in Glottolog
	narratives		52.20724	
Naeini	2 spoken	20.000	32.86043.	cf. Navini (<navinic) in<="" td=""></navinic)>
	narratives	,	53.08112	Glottolog
Yazdi	8 spoken	10.000	31.88138.	cf. Gabri (<zoroastrian td="" yazdi)<=""></zoroastrian>
Zoroastrian	narratives	,	54.37355	in Glottolog
Sivandi	4 spoken	3.000	30.08182.	cf. Sivandi (<central iran<="" td=""></central>
Sivanai	narratives	5,000	52.92264	Kermanic) in Glottolog
Koroshi	7 spoken	10.000	29.52698.	cf. Nourzaei et al. (2015).
	narratives	10,000	52.66519	Balochic in Glottolog
Luri-type	5 spoken	2–3 million	33,50349.	cf. Amān Allāhī Bahārvand
languages	narratives	2 9	49 05946	and Thackston (1986) Luric
languages	hanatives		47103740	in Glottolog
Behbahani	6 spoken	60 000	30 59746	not listed in Glottolog
20112411411	narratives	00,000	50.23304	
Nowdani	2 snoken	3 000	29 79956	not listed in Glottolog
nowualli	narratives	5,000	51 68829	not listed in clottolog
Davani	5 spoken	600	29,70200	cf. Davani (<fars dialects)="" in<="" td=""></fars>
	narratives	500	51 67361	Glottolog
Delvari	2 snoken	4 000	28 76228	not listed in Glottolog
	narratives	.,	51.06716	

Table 7: (continued)

Language	Size of dataset	Approximate n. of speakers	Location	Further information
Dashti	3 spoken narratives	40,000	28.34110, 51.52374	not listed in Glottolog
Lari	4 spoken narratives	50,000	27.67415, 54.32065	cf. Lari (<larestani) in<br="">Glottolog</larestani)>
Bastaki	3 spoken narratives	10,000	27.19754, 54.36614	cf. Bastak (<larestani) in<br="">Glottolog</larestani)>
Bandari	3 spoken narratives	200,000	27.18477, 56.25640	not listed in Glottolog
Minabi	3 spoken narratives	70,000	27.12748, 57.07667	not listed in Glottolog

Table 7: (continued)

References

- Amān Allāhī Bahārvand, Sikandar & Wheeler McIntosh Thackston. 1986. Tales from Luristan (Matalyâ Luriss<u>u</u>): Tales, fables, and folk poetry from the Lur of Bâlâ-Garîva (Harvard Iranian Series 4). Cambridge, Mass: Harvard University Press.
- Belelli, Sara. 2016. A study of language and foklore in the city of Harsin (Kermānshāh province, west Iran). Naples: University of Naples dissertation.
- Brunner, Christopher J. 1977. *A syntax of western Middle Iranian*. Delmar & New York: Caravan Books.
- Cysouw, Michael. 2005. Morphology in the wrong place: A survey of preposed enclitics. In Wolfgang U. Dressler, Dieter Kastovsky, Oskar E. Pfeiffer & Franz Rainer (eds.), *Morphology and its demarcations: Selected papers from the 11th morphology meeting, Vienna, February* 2004, 17–37. Amsterdam & Philadelphia: John Benjamins.
- Dabir-Moghaddam, Mohammad. 2008. On agent clitics in Balochi. In Carina Jahani, Agnes Korn & Titus Paul (eds.), *The Baloch and others. Linguistic, historical and socio-political perspectives on pluralism in Balochistan*, 83–101. Wiesbaden: Reichert.
- Durkin-Meisterernst, Desmond. 2006. *The hymns to the living soul. Middle Persian and Parthian texts in the Turfan collection*. Turnhout, Belgium: Brepols.
- Durkin-Meisterernst, Desmond. 2014. Grammatik des Westmitteliranischen (Parthisch und Mittelpersisch) (Sitzungsberichte Der Phil.-Hist. Klasse 850, Veröffentlichungen Zur Iranistik 73, Grammatica Iranica, Band 1). Wien: Verlag der Österreichischen Akademie der Wissenschaften.
- Erschler, David. 2009. Possession marking in ossetic: Arguing for Caucasian influences. *Linguistic Typology* 13(3). 417–450.
- Gholami, Saloumeh. 2016. Zoroastrians of Iran: Linguistic documentation. *Encyclopedia Iranica*. www.iranicaonline.org/articles/zoroastrians-in-iran-06 (accessed 18 February 2020).
- Gholami, Saloumeh. 2018. Pronominal Clitics in Zoroastrian Dari (Behdînî) of Kerman. In Saloumeh Gholami (ed.), *Endangered Iranian languages*, 111–122. Wiesbaden: Reichert.

- Haig, Geoffrey. 2008. *Alignment change in Iranian languages: A construction grammar approach* (Empirical Approaches to Language Typology 37). Berlin & New York: Mouton de Gruyter.
- Haig, Geoffrey. 2013. The subject/object asymmetry in bound person indexing: Diachronic and discourse considerations. https://www.academia.edu/2759840/The_subject_object_ asymmetry_in_bound_person_indexing_diachronic_and_discourse_considerations (accessed 25 January 2019).
- Haig, Geoffrey. 2017. Deconstructing Iranian ergativity. In Jessica Coon, Diane Messam & Lisa Travis (eds.), *The Oxford handbook of ergativity*, 465–500. Oxford: Oxford University Press.
- Haig, Geoffrey. 2018. The grammaticalization of object pronouns: Why differential object indexing is an attractor state. *Linguistics* 56(4). 781–818.
- Halpern, Aaron. 1998. Clitics. In Andrew Spencer & Arnold M. Zwicky (eds.), The handbook of morphology, 101–122. Oxford: Blackwell.
- Harris, Alice & Lyle Campbell. 1995. *Historical syntax in cross-linguistic perspective*. Cambridge: Cambridge University Press.
- Heston, Wilma Louise. 1976. *Selected problems in fifth to tenth century Iranian syntax*. Pennsylvania: University of Pennsylvania dissertation.
- Ivanow, Wladimir. 1940. The Gabri dialect spoken by the Zoroastrians of Persia. Roma: G. Bardi.
- Jügel, Thomas. 2017. Enclitic pronouns in middle Persian and the placeholder construction. In Mahmoud Jaafari Dehaghi (ed.), Quaterly journal of language and inscription: Dedicated to Professor Mansour Shaki, 41–63. Tehran: Center for the Great Islamic Encyclopeadia. https://www.academia.edu/35077578/Enclitic_Pronouns_in_Middle_Persian_and_the_ Placeholder_Construction (accessed 10 January 2021).
- Jügel, Thomas & Pollet Samvelian. 2016. Les pronoms enclitiques dans les langues ouestiraniennes. *Bulletin de la Société de Linguistique de Paris* 111(1). 391–432.
- Karimi, Yadgar. 2010. Unaccusative transitives and the person-case constraint effects in Kurdish. *Lingua* 120(3). 693–716.
- Korn, Agnes. 2009. Western Iranian pronominal clitics. Orientalia Suecana 58. 159–171.
- Korn, Agnes. 2016. A partial tree of Central Iranian. *Indogermanische Forschungen* 121(1). 401–434.
- Korn, Agnes. 2019. Isoglosses and subdivisions of Iranian. *Journal of Historical Linguistics* 9(2). 239–281.
- Langacker, Ronald. 1977. Syntactic reanalysis. In Charles N. Li (ed.), *Mechanisms of syntactic change*, 57–139. Texas: University of Texas.
- Lecoq, Pierre. 2002. Recherches sur les dialectes Kermaniens (Iran central): Grammaire, textes, traductions et glossaires (Acta Iranica 39). Leuven: Peeters.
- MacKenzie, David N. 1971. A concise Pahlavi dictionary. London: Oxford University Press.
- MacKinnon, Colin. 1977. The New Persian preverb bi-. *Journal of the American Oriental Society* 97(1). 8–26.
- Mahamedi, Hamid. 1982. The story of Rostam and Esfandiyār in an Iranian dialect. *Journal of the American Oriental Society* 102(3). 451–459.
- Mann, Oskar. 1909. *Die Tājīk-Mundarten der Provinz Fārs* (Kurdisch-Persische Forschungen). Berlin: Georg Reimer.
- Menn, Lise & Brian MacWhinney. 1984. The repeated morph constraint: Toward an explanation. Language 60(3). 519–541.
- Mohammadirad, Masoud. 2020. *Pronominal clitics in Western Iranian languages: Description, mapping, and typological implications*. Paris: Sorbonne Nouvelle Paris 3 dissertation.

- Nevis, Joel A. 2000. Clitics. In Geert Booij, Christian Lehmann & Joachim Mugdan (eds.), Morphology: An international handbook on inflection and word-formation, vol. 1, 388–404. Berlin & New York: Mouton de Gruyter.
- Nourzaei, Maryam, Carina Jahani, Erik John Anonby & Abbas Ali Ahangar. 2015. *Koroshi: A corpusbased grammatical description* (Studia Iranica Upsaliensia 13). Uppsala: Uppsala Universitet.
- Öpengin, Ergin. 2013. Clitic/affix interactions. A corpus-based study of person marking in the Mukri variety of Central Kurdish. Paris: Université Sorbonne Nouvelle – Paris 3 dissertation.
- Öpengin, Ergin & Masoud Mohammadirad. Forthcoming. Formal and functional variation in pronominal clitics across Kurdish: Documentation and explanations. In Yaron Matras, Ergin Öpengin & Geoffrey Haig (eds.), *Structural and typological variation in the dialects of Kurdish*. London: Palgrave Macmillan.
- Paul, Ludwig. 1998. The position of Zazaki among West Iranian languages. In
 Nicholas Sims-Williams (ed.), Proceedings of the third European conference of Iranian studies. Part 1: Old and Middle Iranian studies, 163–177. Wiesbaden: Reichert.
- Samvelian, Pollet. 2007a. What Sorani Kurdish absolute prepositions tell us about cliticization. In Frederic Hoyt, Nikki Seifert, Alexandra Teodorescu & Jessica White (eds.), *Texas Linguistic Society IX: The morphosyntax of underrepresented languages*, 263–283. Stanford: CSLI Publications.
- Samvelian, Pollet. 2007b. A lexical account of Sorani Kurdish prepositions. In Stefan Müller (ed.), Proceedings of the 14th international conference on head-driven phrase structure grammar, 235–249. Stanford: CSLI Publications.
- Schmitt, Rüdiger (ed.). 1989. Compendium linguarum Iranicarum. Wiesbaden: Reichert.
- Steele, Susan. 1977. Clisis and diachrony. In Charles N. Li (ed.), *Mechanisms of syntactic change*, 539–579. Texas: University of Texas Press.
- Wanner, Dieter. 1987. *The development of Romance clitic pronouns: From Latin to Old Romance* (Empirical Approaches to Language Typology 3). Berlin & New York: Mouton de Gruyter.
- Windfuhr, Gernot. 2009. Dialectology and topics. In Gernot Windfuhr (ed.), *The Iranian languages*, 5–42. London: Routledge.
- Windfuhr, Gernot. 1975. Isoglosses. A sketch on Persians and Parthians, Kurds and Medes. In Monumentum H. S. Nyberg, vol. 2, 457–472. Téhéran & Liège: Bibliothèque Pahlavi.
- Yip, Moira. 1998. Identity avoidance in phonology and morphology. In Patrick Farrell & Steven G. Lapointe (eds.), *Morphology and its relation to syntax and phonology*, 216–247. Stanford: CSLI Publications.