

Due to the sociolinguistic partitioning of Arabic varieties according to prestige, many grammaticalized forms are often dismissed by most language purists as 'corrupted' or 'deformed', since they deviate in form and meaning from the codified and standardized variety, the use of which garners prestige. The various functional elements examined here, such as the preposition *fī* 'in, at', as an existential particle; possessive linkers, such as *dya*, *taba*, that replaced construct state possessives; the independent pronouns *hu*/*huwa* functioning as interrogatives—are all sociolinguistically marked (i.e. low), despite being natively acquired and used in most informal discourses. The consequence is that constraints are being unduly imposed on the natural evolution of Arabic, which is the official language of more than twenty countries in the Middle East and North African region. Because of the low status assigned to the spoken Arabic dialects by language purists and most native speakers, the aforementioned function words and expressions continued their natural evolution unrestrained, and therefore deviated from the idealized and rigid standard register. Thus, the interface between grammaticalization and sociolinguistics, particularly the attitudes preventing or otherwise promoting change, should be considered in the study of language evolution.

## 4

## Grammaticalization and inflectionalization in Iranian

GEOFFREY HAIG

## 4.1 INTRODUCTION

The Iranian languages constitute a branch of Indo-European, within which their closest relatives are the Indo-Aryan languages. The earliest uncontroversially dated attestation of Iranian is Old Persian, preserved in a series of cuneiform inscriptions located in today's southwestern Iran (the best-known in Behistun), which stem from the 6th to the 4th centuries BCE. In addition to Old Persian, a body of texts known as Avestan (Old and Young) also provide evidence for the oldest layers of Iranian. Avestan texts are ritual and didactic in character, rooted in the belief system of the Zoroastrians. They were 'transmitted orally over centuries and even millennia before being committed to writing some time after 600 CE' (Skjærvø 2009: 45). Thus although the texts clearly represent an ancient form of Iranian (and are very similar to the oldest parts of the Rigveda), the Zoroastrian priests who ultimately committed the Avestan texts to writing spoke much later (Middle Iranian) languages, a fact which has evidently led to some mixing of the linguistic systems that ultimately crystallized in the written form of Avestan, and which makes dating and interpretation of the texts a delicate issue.

Traditionally, Iranian philologists split Iranian into two groups, west Iranian and east Iranian. Although this assumption faces serious (and possibly insoluble) empirical problems, as already pointed out by Sims-Williams (1996), I continue to maintain it here as a pre-theoretical taxonomy. Most of this chapter will deal with what are traditionally termed west Iranian languages. Among the west Iranian languages, Persian has the longest time-depth of attestation, going back to Old Persian of the preceding paragraph, and has enjoyed the greatest cultural and political prestige as the language associated with successive dynasties of Persian empires, and spreading as a language of administration, science, and literature across Asia to the Indian subcontinent. Unsurprisingly, research on grammaticalization, and indeed on historical Iranian morphosyntax in general, has concentrated on Persian. However, the pre-eminence of Persian in the linguistic literature is an artefact of the political and

cultural domination associated with Persian-speakers, rather than reflecting any intrinsically central or salient feature of the Persian language within Iranian. On the contrary, in many respects Persian is an unusual, or even atypical, representative of Iranian (quite comparable to English as a *lingua franca* associated with current economic and political prestige, yet an atypical representative of Germanic). However, the vast majority of other contemporary Iranian languages lack any written attestation beyond a couple of centuries, rendering their historical reconstruction particularly challenging.

The developments in Iranian morphosyntax over the past two millennia exhibit many parallels to the better-known branches of Indo-European, Romance and Germanic. Old Iranian preserved much of the rich (though irregular) inflectional morphology of Proto-Indo-European, including nominal gender, declensional classes, inflectional expressions of aspect, and so on. But the transition from Old to Middle Iranian (around the beginning of the Christian Era) witnessed the collapse and levelling of much of the inherited morphology. These processes are best documented for Persian, which lost gender, all case marking, and entire paradigms in the verbal system (e.g. aorist, old perfect). Morphologically, Middle Persian is thus considerably impoverished when compared to Old Iranian—not unlike the difference between contemporary French and classical Latin. Similar processes of morphological erosion affected most of the other languages, though inherited morphological categories of case and gender have survived in attenuated form in a number of contemporary languages.

Since the great levelling of morphology some two thousand years ago, Iranian languages have been gradually reacquiring morphosyntactic complexity through, for example, univerbation of erstwhile copulas with lexical verbs, the grammaticalization of lexical verbs into modal, aspectual, and voice auxiliaries, the grammaticalization of adpositions to phrasal affixes with case functions, but also in the restructuring and re-creation of person agreement systems on the verbs. Iranian has therefore much to offer for scholars of grammaticalization. However, dedicated research on grammaticalization within Iranian has not yet achieved the same coverage as grammaticalization research in Romance, Slavic, or Germanic (see Davari and Kohan 2017 for recent discussion and references, and Jügel 2015: ch. 2 for the grammaticalization of auxiliary verbs). Given the scale of the issues, the time-depth of attestation, and the number of languages involved, the present chapter is of necessity selective, and will almost exclusively deal with west Iranian. I will begin with a discussion of the grammaticalization—more precisely, the inflectionalization—of person and number agreement from erstwhile pronouns (section 4.2), because this is a topos of grammaticalization research, and the Iranian languages offer an unusually long-term perspective on some of the main issues involved. One of the main findings of this brief survey is that the assumed final stage of grammaticalization, namely into fully-fledged inflection, is an exceedingly slow process, taking millennia before all traces of the lexical, or at least non-inflectional, origins of grammatical formatives are lost. Section 4.3 will take up a variety of other issues in the existing literature, in particular the grammaticalization of auxiliary verbs, and the grammaticalization of case marking. Section 4.4 offers some more general considerations and a summary of the main points.

## 4.2 THE GRAMMATICALIZATION OF PERSON INDEXING IN IRANIAN: SUBJECTS VERSUS OBJECTS

### 4.2.1 FROM PRONOUN TO AGREEMENT AFFIX: THE TYPOLOGICAL PERSPECTIVE

In a number of genetically diverse languages (e.g. Turkic, Bantu), there are striking phonological similarities between free pronouns and the corresponding verbal person agreement affixes. Explanations for these similarities are generally framed in terms of a grammaticalization process: the originally free pronouns have gradually coalesced with verbal hosts, yielding phonologically dependent (clitic or affixal) doublets of the pronouns. These bound forms lose their pronominal status, hence are no longer subject to Binding Conditions, and may co-occur with a co-referential NP in the same local syntactic domain. Ultimately, they become obligatory items of the verb's inflectional morphology: agreement markers. The original pronouns either continue to function as free pronouns or are replaced by innovated pronouns. Siewierska (2004: 251) refers to this cycle as 'a continuous process on-going in all languages in all times'. According to Culbertson (2010), it can currently be observed in spoken French, and the unstressed French subject pronouns *je*, *tu*, etc. can now be analysed as agreement clitics (see Kibrik 2011 for similar claims, and De Cat 2005 for an alternative analysis). The process is often modelled in the form of a cline of form types, as in (1), from Fuss (2005: 4):

- (1) independent pronoun → weak pronoun → clitic pronoun → affixal (agglutinative) agreement marker → fused agreement marker → ∅

Despite different theoretical assumptions, the grammaticalization account of the emergence of agreement has remained dominant in historical linguistics, though with different emphases and terminologies. Indeed, Fuss (2005: 4) refers to the recognition of 'a universal historical pathway' in the rise of agreement markers; see Schnell (to appear) for critical evaluation.

Most previous research has focused on the grammaticalization of subject pronouns. However, object agreement is also cross-linguistically attested, and in the relevant literature it is generally assumed that the grammaticalization of object agreement from object pronouns basically follows the same path as that of subject agreement. Thus Bresnan and Mchombo (1987: 177) claim that the bound object pronouns in Bantu are in the process of grammaticalization into agreement markers, 'parallel to the earlier evolution of the SM [Subject Marker—GH]'. The assumption of a unified grammaticalization pathway for subject and object pronouns has largely remained unchallenged (see van Gelderen 2011 for recent discussion). However, as Siewierska (1999) points out, cross-linguistically, examples of truly obligatory object agreement are vastly less frequent than of subject agreement. If both involve the same mechanisms, it is not readily obvious why this imbalance should obtain.



As it turns out, historical data from Iranian is particularly relevant for this question. In a number of west Iranian languages, an identical paradigm of clitic pronouns came to be used for both subjects, and direct objects, albeit in mutually exclusive domains (with predicates based on past stems and present stems respectively). The roots of this system can be traced back for more than two millennia, so that Iranian provides a natural historical laboratory for tracing the respective developments of subject and object pronouns, each carried by a paradigm of phonologically identical forms. The lesson to be learned from Iranian is that the fates of the two sets of pronouns have been very different: while the subject pronouns have, as predicted by Siewierska's cyclical view, in some languages at least reached the stage of obligatory agreement markers, the object pronouns have basically plateaued at the same stage that obtained more than 2,000 years ago. In what follows I will briefly summarize these developments; see Haig (2018) for more details.

#### 4.2.2 THE CLITIC PRONOUNS OF MIDDLE IRANIAN

The clitic pronouns at the centre of this discussion emerged through syncretism across various non-nominative forms of clitic pronouns, for which cognates are identifiable in Old Iranian and Old Indic (Korn 2009). By the Middle Iranian period some two thousand years ago, these forms had merged to yield a single paradigm of non-nominative clitic pronouns, often referred to in Iranian philology as 'oblique' pronouns. The Middle Iranian forms are provided in Table 4.1, based on Parthian and Middle Persian;<sup>1</sup> the remainder of this section takes up the fate of the cognates of this paradigm in two different functions: transitive subject (A) and direct object (P).

Many west Iranian languages have retained a paradigm of pronominal clitics recognizably cognate with those of Table 4.1. A selection of contemporary West Iranian languages and their pronominal clitics are provided in Table 4.2.

Despite certain minor differences (some involve superficial phonetic processes such as deletion of final *-n*, while others stem from deeper historical origins; see Korn 2009), the overarching similarities across the paradigms are evident, as are the similarities to the Middle Iranian forms of Table 4.1. Not all west Iranian languages have preserved these clitics; some languages of the northwestern peripheries of west

TABLE 4.1. Pronominal clitics in Parthian and Middle Persian

	Singular	Plural
First person	= <i>m</i>	= <i>mān</i>
Second person	= <i>t</i> / = <i>d</i>	= <i>tān</i> / = <i>dān</i>
Third person	= <i>š</i>	= <i>šān</i> / = <i>š</i>

(Jügel 2015: 222)

<sup>1</sup> Table 4.1 ignores some complications, see Skjærvø (2009: 208) for a slightly different version, and Korn (2009) for more detailed presentation.

TABLE 4.2. Pronominal clitics in selected west Iranian languages

	Persian	Mukri (Central Kurdish)*	Hawrami*	Sivand*
1SG	= <i>am</i>	=( <i>i</i> ) <i>m</i>	=( <i>i</i> ) <i>m</i>	= <i>em</i>
2SG	= <i>at</i>	=( <i>i</i> ) <i>t</i>	=( <i>i</i> ) <i>t</i>	= <i>et</i>
3SG	= <i>aš</i>	= <i>ī</i> / = <i>y</i> (after vowels)	=( <i>i</i> ) <i>š</i>	= <i>eš</i>
1PL	= <i>mān</i>	= <i>mān</i> / = <i>in</i>	= <i>mā</i>	= <i>emā</i>
2PL	= <i>tān</i>	= <i>tān</i>	= <i>tā</i>	= <i>etā</i>
3PL	= <i>šān</i>	= <i>yān</i>	= <i>šā</i>	= <i>ešā</i>

\* Sources for the languages other than Persian: Mukri Kurdish: Öpengin (2016: 92); Hawrami: MacKenzie (1966: 25); Sivand dialect: Lecoq (1979: 40). Apparent differences in the qualities of the vowels are in part due to differences in the transcription practices of the sources; they are irrelevant for the present purposes.

Iranian lack them (e.g. Zazaki, Kurmanji Kurdish, Mazanderani), but it is probably reasonable to assume that the earliest stages of these languages also had them. From their earliest attestations, the clitics were used to express adnominal possessors, experiencers, benefactives, and external possessors (see Haig 2008: 105–16), and following the syncretisms among the various non-subject cases, one and the same set was also used for direct objects.

Having dealt with the forms, we turn now to their distribution and functions. In the Old Iranian period, the inherited finite past and perfective verb forms gradually disappeared, echoing similar changes across much of Indo-European, where finite past tense and perfective aspect forms were disappearing. In Iranian, the sole form that remained to effect past tense reference were participles, basically verbal adjectives with resultative semantics (Haig 2008: 41), that had long been in existence in Old Iranian and beyond. In tandem with this change, a fundamental reorganization of the morphosyntax of past-tense clauses occurred, yielding ergative (or non-accusative) alignments in these tenses. These changes have been dealt with in detail elsewhere (Haig 2008: ch. 2; Jügel 2015; Haig 2017) and need not concern us here; for the present purposes it is sufficient to note that in the past tenses of transitive verbs, the paradigm of clitic pronouns introduced in the last section (Table 4.1) also served as transitive subject (A) pronouns (in what follows I will refer to 'subject' pronouns, but in fact only transitive subjects are involved) with all past-tense transitive verbs. Thus the typologically unusual situation arose in which one and the same paradigm of person forms indexed<sup>2</sup> the direct object (P) in present tenses and the A in past tenses (see Arkadiev 2008 for the theoretical implications of this kind of system). The import of this situation for our purposes is that it created a natural laboratory for observing the respective developments of the grammaticalization of subject and

<sup>2</sup> I adopt the term 'indexing' from Haspelmath (2013), as a neutral term for both anaphoric and agreement relations obtaining between a target and controller; the term 'agreement' is reserved for those types of indexing where the presence of the index is obligatory, regardless of presence or absence of the controller in the same clause, and regardless of the information configuration of the entire clause. Agreement is thus a syntactic relation, definable without reference to pragmatics.

object indexing, because both began with phonologically identical input material. It is important to bear in mind that in Old and Middle Iranian, these clitics were special clitics (in the sense of Anderson 2005), whose position was basically after the first constituent of their clause (Wackernagel position).

#### 4.2.3 CLITIC PRONOUNS AS SUBJECTS (A)

In Old Iranian, and well into Middle Iranian, the clitic pronouns used for the A were largely restricted to occurrence in clauses otherwise lacking an overt subject NP. The following examples are from Middle Iranian; (2) shows a clitic pronoun A, while (3) has a NP in the A role, and no clitic pronoun:

- (2) *čē=t ātaxš ī man pus ōzād*  
 because=2SG:A fire of my son extinguish.PST.3SG  
 'because you extinguished the fire of my son [...]' (Middle Persian, Haig 2008: 124)
- (3) *pas ōšbām oy az pidar bōxt [...]*  
 then ōšbām:A 3SG from father rescue.PST.3SG  
 then Ōšbām rescued her from (her) father [...] (Zoroastrian Middle Persian, Jügel 2015: 410, glosses added)

The pronominal nature of these clitics was maintained into the Middle Iranian period. This is evident from the fact that they may be omitted under the condition that their reference is recoverable from the context (see Jügel 2015: 400 for details). The following example shows an overt clitic pronoun for the A of the first clause, and zero for the co-referential A of the subsequent clause:

- (4) a. *u=š ardawān ōzad [...]*  
 and=3SG:A Ardawān kill.PST.3SG  
 b. *ud duxt ī ardawān pad zanīh kard*  
 and daughter of Ardawān to wife make.PST.3SG  
 'And he<sub>i</sub> killed Ardawān [...] and (he<sub>i</sub>) took his daughter as wife'  
 (Zoroastrian Middle Persian, Jügel 2015: 411, glosses added)

From Jügel (2015), two facts emerge that support the interpretation of the Middle Iranian clitic A-pronouns as pronouns, rather than agreement: first, the incompatibility of the pronoun with a free expression of the A,<sup>3</sup> and second, the ability to be omitted under pragmatically felicitous conditions, in a manner comparable to the omission of a free pronoun. We will follow this analysis for the time being, but in the following discussion I will suggest that this is probably not the whole story.

The system of indexing the A through a pronominal clitic has disappeared in some west Iranian languages, notably Persian, but elsewhere it has survived remarkably

<sup>3</sup> Jügel (2015: 396–9) discusses the few examples from Middle Persian where the A-clitic is doubled by an overt A in the clause (94 attested in a corpus of 6,815 clauses). Most of these involve some form of dislocation, e.g. an afterthought or a preposed constituent, with unclear clausal loyalties, or are scribal errors. Otherwise, clitic pronoun and overt subject NP are mutually exclusive.

well. In Central Kurdish, the system is still recognizably that of Middle Iranian, but with one very crucial difference: the pronominal clitics that index an A in the past tenses have become fully obligatory: 'every single past transitive construction requires an A-past clitic', regardless of the presence or absence of an overt A constituent in the same clause (Haig 2008: 288), or any other pragmatic conditions. The following examples illustrate the co-occurrence of subject NP and clitic in Central Kurdish, and in two other Iranian languages with a similar system:

- (5) *eme to=mān nard bo šar-ī*  
 1PL:A 2S:P=1PL:A send:PST to city-OBL  
 'We sent you to the city' (Central Kurdish, Mukri dialect, Ergin Öpengin p.c.)
- (6) *me ketaw=em xeri*  
 1SG book=1SG.A buy.PST.3SG  
 'I bought the book' (Laki, Dabir-Moghaddam 2008: 96)
- (7) *mæ=m ketav ese*  
 1SG=1SG.A book buy.PST.3SG  
 'I bought the book' (Davani, Dabir-Moghaddam 2008: 93)

In Central Kurdish, Laki, and Davani, we have a typologically unusual kind of subject agreement, in which the subject index itself is a mobile clitic, clearly reflecting its pronominal origins. But unlike a typical subject pronoun, the clitics of Central Kurdish are not omissible, even in environments where pronouns are normally dispreferred or even disallowed (e.g. same-subject clause coordination, or subject relativization). Conditions of space preclude illustration of these properties; I refer to the ample documentation of Central Kurdish subject clitics in MacKenzie (1961, 1962) and Öpengin (2016); all recent research converges on the verdict that they are exponents of an agreement relationship (Samvelian 2007; Haig 2008; Öpengin 2016).

A further stage in the assumed grammaticalization of subject agreement is attested in other west Iranian languages, for example the dialect of Semnan (Majidi 1980). In the past tenses, we find two distinct paradigms of person agreement suffixes on the verb, one for transitive verbs (in bold type) and the other for intransitives, illustrated in Table 4.3 (from Majidi 1980: 119, transcription and segmentation adapted; the initial prefix is an indicative marker). For ease of comparison, the pronominal clitics of Middle Iranian from Table 4.1 have been added to Table 4.3.

It is evident that in Semnan, the paradigm of person suffixes for transitive verbs is distinct from that of intransitive verbs. Furthermore, we can assume that the paradigm found with transitives is an innovation, and its source is the clitic pronoun paradigm of Middle Iranian, as shown in the right-hand column. Presumably the Semnan system arose when the erstwhile clitic pronouns lost their syntactic mobility, and grammaticalized to the verb, becoming effectively inflectional agreement morphology.<sup>4</sup> Systems like

<sup>4</sup> My account of Semnan is based on Majidi (1980). However, Masoud Mohammadirad (p.c.) reports that the subject-indexing clitics on the past transitive verb are not obligatory (based on recent fieldwork with speakers) in Semnani, and are sometimes omitted. If this is confirmed, then the account provided here needs to be modified accordingly. Such a state of affairs would actually provide further support for the fact



TABLE 4.3. Intransitive and transitive person indexing (past indicative), Semnan dialect

	Intransitive 'die'	Transitive 'do, make'	Middle Iranian clitic pronouns
1SG	<i>ba-mard-un</i>	<i>hā-kard-an</i>	= <i>m</i>
2SG	<i>ba-mard-e</i>	<i>hā-kard-at</i>	= <i>t</i> / = <i>d</i>
3SG	<i>ba-mard-e</i>	<i>hā-kard-eš</i>	= <i>š</i>
1PL	<i>ba-mard-in</i>	<i>hā-kar-mun</i>	= <i>mān</i>
2PL	<i>ba-mard-in</i>	<i>hā-kar-tun</i>	= <i>tān</i> / = <i>dān</i>
3PL	<i>ba-mard-an</i>	<i>hā-kar-šun</i>	= <i>šān</i> / = <i>š</i>

the Semnan one are also noted in other Iranian languages (Jügel 2015: 463–4). It should also be noted that even in those languages like Central Kurdish, or Gorani, where the subject index retains its syntactic mobility, the verb itself is a frequent host option for the clitic; cf. Gorani *kamtar ward=iš?* (vulture ate.PST=3SG.A) 'did a vulture eat (them)?', where the third singular subject index attaches to the verb *ward* 'ate' (Mahmoudveysi, Bailey, Paul, and Haig 2012: text 2:55). Thus the development in Semnan, where the erstwhile clitic is now exclusively found on the verb itself, represents the grammaticalization of an already available positional variant, rather than a completely novel development.

Schematically, the development of subject agreement from clitic pronouns in Iranian can be sketched as in Table 4.4. Here I distinguish the dimensions of 'obligatoriness', or 'inflectionalization', from the degree of phonological bondedness (Kibrik 2011; Norde 2009). The latter subsumes two criteria: morphological integration into the host and degree of freedom of host selection.

The sequence of stages set out in Table 4.4 presupposes that the predecessors of Central Kurdish and Semnan dialect had a clitic system similar to that of the attested Middle Iranian languages Parthian, Middle Persian, and Bactrian.<sup>5</sup> As we have no records of the immediate predecessors of Kurdish and Semnan, this is obviously hypothetical, but it nevertheless appears to be plausible in the light of what is known about the pronominal clitic system across west Iranian. Note also that Table 4.4 represents but one possible line of development. Others are attested, most notably for Modern Persian, where the pronominal clitics simply disappeared from the subject function, and Persian past-transitive verbs came to carry subject agreement

that the form of the index is not necessarily a good predictor of its function. Although the subject index in Semnani looks superficially like an agreement affix (i.e. fixed position on the verb), it has apparently retained some of its pronominal properties. Similar mismatches between form and function characterize the object indexes which are discussed below.

<sup>5</sup> It is important to note that we lack direct records for the predecessors of these contemporary languages. In fact they cannot be direct descendants of either Middle Persian or Parthian, because we assume that common Kurdish must have preserved the inherited Iranian case distinctions, and Semnani does so to this day. In Parthian and Middle Persian, however, case had already been largely lost. Central Kurdish, Semnan, etc. must therefore go back to some contemporary of Parthian etc., which had retained case (and gender), but for which no historical records exist.

TABLE 4.4. Schematic summary of the inflectionalization of person indexing in Iranian

Stage/Language	Syntactic mobility ('bondedness')	Degree of inflectionalization
Old Iranian/Middle Iranian transition Verb forms (participles) lacking subject-indexing morphology become the norm for past transitive constructions; use of genitive/dative clitic pronouns as 'subjects' begins, probably inherited from existing non-canonical subject constructions (Haig 2008).	Syntactically mobile, high freedom of host selection, Wackernagel position.	Not obligatory, clearly still pronominal; absolute numbers of relevant examples in the extant corpus is too limited to draw firm conclusions.
Middle Iranian (Middle Persian, Parthian, Bactrian, Jügel 2015) Participles now the norm for past-tense transitives, generalized 'oblique' pronoun is normal expression of subject.	Syntactically mobile, Wackernagel position. High freedom of host selection (including subject NP itself, or complementizer).*	Not obligatory, but already significantly more frequent than would be expected of a corresponding free pronoun (see below in this section).
Contemporary Central Kurdish (e.g. Öpengin 2016)	Syntactically mobile, VP-based position, some freedom of host selection, but subject NP and complementizers are no longer possible hosts. Morphological integration into predicate is possible.	Obligatory agreement marker (but see n. 4).
Contemporary Semnan dialect (Majidi 1980)	Positionally fixed (bonded to verb stem).	Obligatory agreement marker.

\* Jügel (2015: 249) finds 'no restrictions [*keine Einschränkungen*, GH]' on the host category for the Middle Iranian clitic pronouns, except for prepositions.

morphology from a different source (identical to that used for intransitive verbs, suggesting that the latter is the source candidate). The other possibility is that the non-obligatory nature of the Middle Iranian system continued, which seems to be the case in Taleshi (Paul 2011). However, a system with obligatory subject indexing (agreement), still carried by mobile clitic pronouns, is certainly a possible outcome in the west Iranian context. The Semnan dialect shows how these pronouns may finally lose their syntactic mobility and become in effect verbal affixes, part of verbal inflection.

A question that needs to be addressed is the nature of the mechanisms involved in the shift from clitic pronoun to agreement, and the chronology of the events. Thanks to Jügel's (2015) rich documentation of the Middle Iranian facts, it is now possible to advance at least some tentative hypotheses in this direction. As already mentioned, Jügel himself analyses the mobile subject clitics of Middle Iranian as pronouns, mainly due to the lack of attested cases of clitic doubling, which appears to be his main diagnostic for distinguishing pronoun from agreement marker (cf. Jügel 2015: 267–8, n. 638). On this view, the clitic subject pronouns of Middle Iranian are basically prosodically deficient versions of normal pronouns. However, Jügel's own material actually suggests that this is not the whole story. The most striking piece of evidence is the sheer frequency of occurrence of the clitic pronouns, which can be inferred from Jügel (2015: 326, table 5.4): in the largest corpus, Middle Persian, around 44 per cent of all past transitive clauses contained a clitic pronoun exponent of the subject (N=6,815). The full significance of this figure emerges when we compare it with the percentage of overt pronouns in transitive clauses of other languages which allow referential null subjects. Relevant figures for contemporary spoken Persian (Adibifar 2016), Cypriot Greek (Hadjidas and Vollmer 2016), and Northern Kurdish (Haig and Thiele 2016) are: Persian: 8 per cent (N=603), Cypriot Greek: 4 per cent (N=494); Northern Kurdish: 29 per cent (N=422).<sup>6</sup> For these languages, and indeed most others that allow null referential subjects, the favoured form of expression for subjects is zero, not pronominal. The figures from Middle Persian are thus distinctly odd, and require explanation. I would tentatively suggest that the clitic pronouns had already at the Middle Persian stage undergone a rapid rise in frequency, and that this frequency increase should be seen as the precursor to the later grammaticalization (cf. Meyerhoff 2000 for a frequency-based account of the grammaticalization of subject agreement, and Schnell, to appear, for critical discussion).

Jügel cautiously relates the development towards the agreement system of Central Kurdish to the presence of 'topic agreement' (Jügel 2015: 463–4), i.e. the use of a resumptive clitic pronoun following a stage-setting, left-dislocated new topic, which is attested in some Middle Iranian texts. The construction would, on this account, have lost its pragmatic markedness after the Middle Iranian period, and become reanalysed as subject–verb agreement in Central Kurdish, and other languages with this kind of agreement (basically in line with Givón's 1976 account of the emergence of agreement via topicalization of pronouns). But while these structures may have played some role in the process, they cannot account for the overall jump in

<sup>6</sup> The comparatively high figure for Northern Kurdish is noteworthy; it is probably related to the fact that in the Northern Kurdish corpus, many of the verbs are past-tense transitives, which lack overt agreement morphology. It is likely that in these contexts, more overt pronouns are used as compensation for the lack of overt subject indexing, as proposed by Kibrik (2011) with reference to Russian past-tense forms (likewise lacking person agreement). This needs more research; it also underscores the inapplicability of categorical typologies along the lines of 'pro-drop' vs 'non pro-drop'; there are evidently intermediate degrees, reflected in graded frequency of omission in discourse.

frequency in the Middle Iranian data. In the light of cross-linguistic research on pronoun omission in transitive clauses, the figure of 44 per cent pronoun retention is highly significant, and indicates that these clitic pronouns were qualitatively distinct from free pronouns; if this is on the right track, then the grammaticalization process from pronoun to agreement marker was already well under way at the Middle Persian stage.

These considerations underscore the necessity to look beyond the issue of 'clitic doubling' as a diagnostic for the degree of grammaticalization, and to take comparative-frequency data into account. What makes a pronoun a pronoun is not just its inability to appear in the same local domain as its antecedent (lack of 'clitic doubling'). Pronouns are also characteristically prone to be omitted under conditions of pragmatic identifiability of the referent. Indeed, many of the Jügel's Middle Iranian examples illustrate clitic pronouns in contexts where pronouns would not normally be expected, for example the following:

- (8) *ēk, ke=š maṇ brēhēnīd*  
 one, that=3SG.A 1SG create.PST.3SG  
 'one which created me' (lit. 'one that he created me', Zoroastrian Middle Persian, Jügel 2015: 378)

The subject pronoun =š attaches to the complementizer/relativizer, although resumptive pronouns are generally not required in Iranian subject relativization. The suggestion I am making is that the grammaticalization process begins with an extension of the clitic pronouns into syntactic environments where pronoun omission (zero) would generally be the norm. Note that such an extension will not necessarily yield instances of clitic doubling; it will simply yield an overall statistical increase in the frequency of overt pronouns, and a corresponding drop in the rates of zero anaphora.

Available cross-corpus research indicates that in the case of transitive clauses, somewhere between 75 per cent and 95 per cent of subjects have non-lexical expressions, i.e. are either pronominal or zero (see Haig and Schnell 2016). Throughout its attested history, Persian is known to be a language that licenses referential null subjects and, concomitantly, avoids pronominal subjects. Thus the most common form for transitive subjects in Persian is zero, and there is little reason to suppose that this has changed significantly over the history of the language. Against that background, the frequency data extractable from Jügel (2015) is extremely revealing for understanding the grammaticalization process. For the time being, I conclude that the clitic subject pronouns of Middle Iranian, while not agreement markers in a strict sense, nevertheless differed in their distribution significantly from free subject pronouns in other Iranian languages (and, I suspect, from free pronouns in present-tense transitives in Middle Iranian, though this awaits further research). Kibrik's (2011) notion of 'bound tenacious pronoun', implying a prosodically bound form that is approaching (to different degrees) agreement status, would perhaps be appropriate. Having outlined the development from clitic pronoun to verbal agreement affix, we now turn to consider the fate of the cognate clitic pronouns in object functions.



## 4.2.4 PRONOMINAL CLITICS AS OBJECTS

The development of clitic pronouns in the object role is simpler than that just sketched for the subject role, and this section is correspondingly brief (see Haig, 2018, a for a more detailed discussion). Clitic pronouns in the direct object function are attested in earliest records of Iranian, so we may assume that this was a syntactic possibility available perhaps for as long as 3,000 years. In Old Iranian, there was still a dedicated paradigm of accusative clitic pronouns, which later syncretized with the other non-nominative clitic pronouns to yield the paradigm provided in Table 4.1. Examples from Old Persian, with the still-distinct form of the accusative pronoun, are the following (note again the Wackernagel position of the clitic):

- (9) *pasāva=dim manā frābara*  
 after.that=3SG.ACC 1SG.GEN/DAT bestow.PST.3SG  
 'After that he bestowed it on me' (Old Persian, Haig 2008: 47)
- (10) *kāra hya aθuriy hau=dim abar yātā Bābirauv*  
 people which Assyrian, DEM=3SG.ACC bring.PST to Babylon  
 'The Assyrian people - they brought it to Babylon' (Old Persian, Haig 2008: 47)

Examples of Middle Iranian clitic pronouns in object function are given below (from Haig 2008: 115):

- (11) *čīd=mān pāyēd*  
 always=1PL:P protect.PRS.3SG  
 '(It) always protects us'
- (12) [...] *u=š hamēw bōžēnd*  
 [...] and=3SG:P always save.PRS.3PL  
 '(the Gods) always save him'

Throughout Old and Middle Iranian, pronominal clitics could only express the object in the absence of an overt object NP. Thus the object clitics were, despite their clitic status, fairly obviously pronouns, rather than any kind of agreement: they were in complementary distribution to a free NP (or full-form pronoun) object (Jügel 2015: 399). Some 1,500 years later, the reflexes of the Middle Iranian clitic pronouns remain widely attested as clitic object pronouns in numerous contemporary west Iranian languages. The main change has been what Haig (2008) refers to as 'rightward drift' with regard to clitic placement: the Old and Middle Iranian Wackernagel position has given way to a VP-based clitic placement system, with the finite verb itself as a very common host (Wackernagel position is retained in a small number of languages). Typical examples of clitic pronouns attaching to the finite verb are the following:

- (13) *hālā ne-mi-bin-am=aš*  
 now NEG-IND-see.PRS-1S=3SG:P  
 'now I don't see it' (Modern Persian, Roberts 2009: 256)

- (14) *m-war-im=šān*  
 IND-eat.PRS-1SG=3PL:P  
 'I will eat them' (Gorani, Mahmoudveysi et al. 2012: 67)
- (15) *sob mo-gor-im=eš*  
 morning IND-take.PRS-1PL=3SG:P  
 'In the morning we will take it' (dialect of Sivand, Lecoq 1979: 91)

In some languages, the clitics also occur as what are arguably 'endoclititics', i.e. enveloped within inflectional verbal morphology, most notably in Central Kurdish (see Harris 2002 on endoclitisis, and Öpengin 2016 for analysis of the relevant facts for Kurdish). The following examples illustrate the position of the object clitics in the Mukri dialect of Central Kurdish (Northwest Iranian, West Iran, Öpengin 2016):

- (16) a. *kut=i segbāb bo de=m=guž-i?*  
 say.PST=3SG.A dog.son why IND=1SG.P=kill.PRS-2SG  
 'He said: 'Son of a dog, why are you killing me?'
- b. *kut=im bāb=im nā=t=guž-im*  
 say.PST=3SG.A brother=POSS1SG NEG=2SG.P=kill.PRS-1SG  
 'I said: 'O brother, I am not killing you'' (Öpengin 2016, ŽB 183-4)

Evidently the 'clitics' here resemble affixes, because they are morphologically integrated into the predicate. Functionally, however, the object clitic of Central Kurdish continues to be in complementary distribution with a NP object. In other words, the presence of the clitic pronoun is only licensed in the absence of the coreferential object. This is demonstrated in (17), where both direct objects are expressed as free pronouns (in bold type), and no corresponding clitic pronoun is allowed on the verb:

- (17) *emin de-kāte weziri desti.řāstī šā fēbās-i w*  
 1SG(P) IND-make.PRS.3SG Vizier-of right.hand.of Shah Abbas-OBL and  
*eto de-kāte kālek-firoš*  
 2SG(P) IND-make.PRS.3SG melon-seller  
 '(God) is appointing (lit. making) **me** the right-hand vizier of Shah Abbas and making **you** a melon-seller.' (Öpengin 2016, KF.118-19)

In sum, the history of clitic object pronouns in Iranian can be reliably traced back for over 2,500 years. In terms of the formal properties of the exponents, from the earliest attestation they exhibit typical properties of special clitics, such as syntactic mobility (clause-second in Old and Middle Iranian, VP-second in much of Kurdish) and freedom of host selection, though in some languages they may also be closely integrated into the predicate, as in Central Kurdish. Yet despite their lack of prosodic independence, they have generally failed to advance down the postulated grammaticalization cline beyond the stage of pronouns. In other words, throughout the entire attested history of west Iranian, the object clitics have not evolved into obligatory object agreement in the category of person in any language known to me (see Jügel 2015: 463, n. 1032). Instead, the clitic pronouns have remained pronominal expressions of the object, in complementary distribution with co-referent-free NP objects, and also

omissible if the object is pragmatically recoverable. With regard to the proposed grammaticalization cline from pronoun to agreement (1), they have basically remained stuck at the same stage for 2,500 years, namely as as clitic pronouns.<sup>7</sup>

#### 4.2.5 SUMMARY: THE (NON-)GRAMMATICALIZATION OF IRANIAN CLITIC PRONOUNS

The developments of the Iranian clitic pronouns provide us with a natural laboratory for investigating the differences between subject and object grammaticalization processes, because the initial phonological material was identical for both (cf. Table 4.1), and both began their careers as Wackernagel clitics. In their later developments, however, they have diverged remarkably. The object clitic pronouns have basically remained just that: prosodically dependent object pronouns, in complementary distribution with free-form objects. Nowhere can we find a convincing case that they have shifted closer towards an agreement system, bar sporadic cases of clitic doubling mentioned in n. 7. Note however that this statement applies to object agreement expressed by the descendants of the clitic pronouns. Object agreement is possible in the category of gender (e.g. in Tati: see Stilo, to appear). But the exponents of gender agreement do not originate in pronominal clitics, and thus are the outcome of a different process from the one discussed here. As for the clitic pronouns used for transitive subjects, there was indeed a shift from alternating to obligatory, precisely in line with the predictions of grammaticalization theory.

#### 4.3 FURTHER TOPICS IN IRANIAN GRAMMATICALIZATION

In this section I will survey some current topics in grammaticalization within Iranian, with particular emphasis on findings of high typological relevance. Section 4.3.1 revisits a classic of Iranian grammaticalization, the object marker *=râ*, while 4.3.2 looks at a less well-known instance of the grammaticalization of an auxiliary.

##### 4.3.1 THE GRAMMATICALIZATION OF DIRECT OBJECT CASE MARKERS

A well-known example of grammaticalization in Iranian is that of object case markers. As mentioned, already by Middle Iranian, some west Iranian languages had lost the

<sup>7</sup> In fact, 'colloquial registers' of Persian do permit sporadic instances of object clitic doubling (Samvelian and Tseng 2010), and this has been argued to represent an instance of agreement. However, the cited examples are pragmatically quite marked. Such clitic doubling is not possible with e.g. indefinite or focal controllers (Rasekh 2014). I am not aware of any pragmatically neutral syntactic configuration in Persian where object clitic doubling is obligatory. This would seem to preclude an analysis as agreement in the narrower sense. Of course, different researchers delineate 'agreement' in different ways (see Corbett 2006), and the differing viewpoints are in part terminological in nature.

Old Iranian case morphology, meaning that with the exception of some minimal vestiges in the pronoun system and in kinship terms, Middle Persian and Parthian no longer marked direct objects (Jügel 2015: 192). Modern Persian, however, has renovated its case system, and now regularly marks specific<sup>8</sup> direct objects via a clitic [=râ:], often realized as [=rɔ:], or just [=ɔ:]. Different sources use different conventions for representing this clitic; I render it orthographically with *=râ*, regardless of the source, or degree of phonetic attrition. Non-specific direct objects remain unmarked (i.e. modern Persian has DOM). A simple example is the following:

- (18) *Sârâ=râ did-am*  
Sara=ACC see.PST-1SG  
'I saw Sara' (Bohnacker and Mohammadi 2012: 62)

*Sârâ*, as an inherently definite proper noun, requires overt object marking. The development of the Persian object marker was identified by Hopper and Traugott (1993: 157–60) as an example of the development of the cline shown in (19):

- (19) lexical word > postposition > suffix

Note, however, that contemporary *=râ* is at best a phrasal affix, rather than a component of nominal morphology; it attaches to the final element of its NP, which might, for example, be an adjective:

- (20) *lebâs-e sefid=râ xarid-am*  
dress-of white=ACC buy.PST-1SG  
'I bought the white dress'

In terms of stress placement, Kahnemuyipour (2003) notes that *=râ* is non-stress bearing, and is 'outside the phonological word' (p. 339). Thus we need to interpret the 'affix' stage of the cline in (19) fairly loosely, to include clitics and phrasal affixes; even after a millennium of attestation as an object marker, *=râ* is not a fully morphologically integrated affixal case marker of the kind that characterized the Old Iranian case system.

The roots of this accusative clitic can be traced back ultimately to a nominal element *râdiy*, meaning something like 'because, on account of', already used in Old Persian as a postposition in this sense (Kent 1953: 205; Paul 2017), requiring the genitive case of its complement. From this, a postposition *rây* developed, with benefactive, possessive, and recipient semantics in Middle Iranian, which continued into Early New Persian (approx. 900–1100 CE). But already in Middle Persian, it had become extended to use as a marker of definite direct objects, though the exact pathway of the development remains obscure. A Middle Persian example with a direct object is the following:

<sup>8</sup> 'Specific' is to be understood here as shorthand for 'related to a complex bundle of factors involving pragmatic identifiability, topicality, and specificity'. In fact, the factors determining DOM in Persian remain disputed; see Paul (2008) for discussion and references.



- (21) *ka ān šagrān rāy zīndag ō amā āwarēd*  
 that those lions-PL ACC living to 1PL bring.2PL  
 'That you bring those lions to us alive' (Jügel 2015: 214)

The semantic pathway transcribed by *rādiy* is thus one of lexical to grammatical; from a more concrete lexical meaning 'because, on account of', to marking a particular grammatical relation, that of direct object (in fact, with no clear semantic core). Traces of its origins remain, however, for example in the interrogative pronoun *čērā* (what=*rā*), lit. 'for what, why'. The benefactive/possessive meaning of *=rā* survived as late as the 19th century CE, as in the following:

- (22) *ahālī-ye orūpā=rā 'aqīde īn ast ke*  
 people-of Europ=*rā* opinion this is that  
 'The people of Europe are of the opinion that [...]'  
 (lit. 'to/for the people of Europe the opinion is ...', Paul 2008: 335)

Similarly, *=rā* continues to mark left-dislocated, frame-setting topics:

- (23) *in dar=rā, qofl=aš=rā diruz šekast-am*  
 this door=*rā*, lock=POSS.3SG=*rā* yesterday break.PST=1SG  
 'This door, I broke the (lit. its) lock yesterday' (p.c. Mohammad Rasekh-Mahand)

Remnant semantic content is also visible in the restriction on marking only specific direct objects, possibly an inheritance of its origin as a marker of recipients, which in discourse tend to be overwhelmingly definite. However, one would, on this view, also expect the feature of +/- human to be relevant in modern Persian DOM, but this does not appear to be the case. Hopper and Traugott (1993: 159–60) consider the shift towards direct object marking to involve a 'contraction of range with respect to thematic roles', but this view is based on the assumption that the direct object role is restricted to Patients and Themes in Persian, which is not the case (the object in (18), for example, is not a Patient). In fact, the semantic development fits well with the most basic assumption of grammaticalization as involving a loss of lexical meaning; the direct object role is defined syntactically, not semantically, and the association of *=rā* with this role (though the match is not perfect, as shown above) can indeed be interpreted as a clear case of a shift from semantic to grammatical function.

It is worth pointing out that there is nothing inevitable in the Persian developments just sketched. Other west Iranian languages likewise lost inherited case morphology, but have not to this day replaced it (e.g. Central Kurdish, Southern Kurdish), leaving subjects and objects equally unmarked. Notably, there has been no shift towards (S)VO in these languages, which remain, as does the totality of Iranian, (S)OV. Elsewhere, innovated object markers have been recruited, though from sources etymologically distinct from the *rādiy* postposition discussed above (see Windfuhr 1992; Haig 2008: ch. 4; Stilo 2009; Paul 2017 for discussion of Iranian case systems), while in other contemporary languages, a particle cognate with *rā* is attested, but it has not become an accusative marker. Thus the grammaticalization of *rādiy*>*rā* in Persian emerged through a contingent combination of factors that together yielded this specific development. The larger pan-Iranian framework, however, is the renewal

of strategies for object marking which has led to several distinct solutions in the individual languages.

#### 4.3.2 THE GRAMMATICALIZATION OF AUXILIARIES

Perhaps the most fertile area of grammaticalization in Iranian involves the renewal of verbal TAM categories from erstwhile full verbs, a topic that has also been central to grammaticalization research since its inception (see Hopper and Traugott 1993: 42–5 for early discussion, Hengeveld 2011 and Narrog 2012 for recent developments). Well-known cases include the development of future tense markers from verbs of motion (English *gonna* < *going* to, Spanish *ir* 'go' + *a* + infinitive), or from 'have' in Romance. A particularly well-documented case in Iranian is the development of an analytical future tense with an auxiliary verb originally meaning 'want', *xāstan*. The development of a future marker from 'want' has obvious parallels in, for example, Germanic, and I will not discuss the Persian case here. For the grammaticalization and univerbation of 'be', see Jügel (2015: 123–49). Other cases of grammaticalization involve the modern Persian modal particle *bāyad* (obligation), from an erstwhile finite verb construction, the development of a passive auxiliary from 'come' (e.g. *hatin* 'come') in Kurmanji Kurdish (Öpengin and Haig 2014), or from 'become', as in Persian *šodan*, which itself goes back to a verb of motion, cf. Old Persian *š(i)yav-* 'to set, go forth' (Cheung 2007: 40). A development that can be related to the grammaticalization of auxiliaries is the emergence of complex predicates, consisting of a non-verbal element plus a light verb, to express numerous basic verbal meanings (basically, this is the main strategy for creating new verbal lexemes in much of Iranian, where productive derivational morphology for this function is lacking; see e.g. Haig 2002 and Samvelian and Faghiri 2013, for discussion of complex predicates in Iranian). Finally, I should point to a typologically very rare type of grammaticalization that has recently been discussed in Iranian linguistics, namely the development of definiteness suffixes (e.g. in Central Kurdish) from diminutives; see Jahani (2015) and Haig (to appear) for discussion.

A less well-described and cross-linguistically more unusual development is the grammaticalization of a continuous marker from a lexical 'have' verb. A verb 'have' is not mentioned as a lexical source of continuous aspect in Heine and Kuteva (2002), so we can assume that the development is fairly unusual. In this section I will briefly outline the facts, while referring to Davari and Kohan (2017) for a more detailed exposition.

#### 4.3.3 CONTINUOUS ASPECT FROM 'HAVE': COLLOQUIAL PERSIAN DĀŠTAN

Persian verbs have two stems, called here a past and a present stem, respectively associated (approximately) with past and present temporal reference. Both stems form the base for a number of distinct TAM formations. The present tense is broadly characterized by an opposition between forms prefixed with a stressed prefix *mi-*, used for all forms of the indicative, and forms that lack this prefix. The latter may be

TABLE 4.5. Preliminaries for the Persian tense and modality system

Gloss	Past stem	Present stem	ind.prs.1sg	sbj.prs.1sg
do, make	<i>kard</i>	<i>kon-</i>	<i>mi-kon-am</i>	<i>be-kon-am</i>
hit	<i>zad</i>	<i>zan-</i>	<i>mi-zan-am</i>	<i>be-zan-am</i>
go	<i>raft</i>	<i>rav-</i>	<i>mi-rav-am</i>	<i>be-rav-am</i>

prefixed with *be-* (with variant *bo-*), generally referred to as a subjunctive prefix, or may lack a prefix entirely (as in certain imperatives). The basics of this system are illustrated in Table 4.5 for three common verbs.

Within this system (largely adhered to in the standard written language), there is no grammaticalized distinction of aspect in the present tense. In other words, the same indicative present verb form with *mi-* may be used for habitual, for punctual, or for ongoing and continuous events; aspectual nuances are supplied contextually, or must be inferred from the lexical verb semantics (Aktionsart), as illustrated in the following (*mi-* with a present tense verb is glossed with INDIC(ATIVE):<sup>9</sup>

- (24) *dar bâreye mosabeq-e futbol sohbat mi-kon-im*  
 about game-EZ football conversation IND-DO.PRS-1PL  
 'We are talking about the football game.'  
 (process, interpretable as referring to the current time of speech)
- (25) *mâ dar kelâs ingilisi sohbat mi-kon-im*  
 1PL in class English conversation IND-DO.PRS-1PL  
 'We (generally) talk English in the classroom.'  
 (process, interpretable as referring to a habitual event, rather than the time of speech)
- (26) *vaqti suzan=râ dar bâdkonak foru ø-kon-i*  
 when needle=ACC in balloon downward SBJ-DO.PRS-2SG  
*mi-tark-ad*  
 IND-burst.PRS-3SG  
 'If you push a needle into a balloon, it bursts'  
 (punctual, appropriate for referring to a general truth holding at all times)
- (27) *man esm=aš=râ mi-dân-am*  
 1SG name=POSS.3SG=ACC IND-know.PRS-1SG  
 'I know his name'  
 (state, obtains for an unbounded period, including the time of speech)

The west Iranian verbal system underwent a transition from the aspect-based system of Old Iranian to a tense-based system, built on the fundamental opposition between

the two stems. In the wake of this major restructuring, a number of contemporary Iranian languages have since developed secondary aspectual distinctions, recruited from various sources. The *mi-* prefix just discussed is the result of such a grammaticalization, going back to an erstwhile adverb, *hamê* 'always' (Windfuhr 2009: 24). There are numerous functionally parallel (though etymologically distinct) counterparts across Iranian, yielding a particularly compelling case of parallel grammaticalization across related languages (or 'drift', to use Sapir's term); an example is the *di-* prefix in Northern Kurdish, illustrated in (36) and (37).

Although I have argued here for an aspect-neutral analysis of the modern *mi-* prefix with present stems, its origins as an aspectual marker are still evident from two facts: First, it does not occur with the inherently state predicate *dâstan* 'have' (present stem *dâr-*), or with the defective copular verb *hastan* 'be'. Thus to express indicative present 'I have' the expected form *\*mi-dâr-am* is not possible, and instead we find *dâr-am*. This is presumably due to an incompatibility of the original continuous meaning of the source of *mi-* with the stative meaning of *dâstan*, or with the copula, quite comparable to the ungrammaticality of present continuous with English *have* (cf. *\*I am having* to express a state of possession). Otherwise, however, *mi-* is compatible, and indeed obligatory, with indicative present forms for all Persian verbs, regardless of their aspectual semantics, and irrespective of whether they are negated or interrogative. Second, the aspectual component of *mi-* is preserved with past stems, where it optionally expresses imperfective (past continuous, but also certain modal nuances), and forms an opposition with an unmarked perfective past. In combination with the present stem, however, this opposition is simply not available (in the morphology at least); hence I gloss *mi-* in the present as IND,<sup>10</sup> but in the past as IMPV.

Where these erstwhile aspect markers have become bleached of aspectual content (as I have argued above for Persian *mi-*), additional aspectual markers may be added to the system. In spoken Persian, an innovated continuous form has emerged, based on a finite form of the verb *dâstan* 'have'. In this construction, the main verb is likewise finite; it must take the same tense as *dâstan*, carry the *mi-* prefix, and the appropriate person and number agreement. The first two examples illustrate the past tense, while (30) and (31) illustrate the present tense (the forms of *dâstan* are glossed (PROG)RESSIVE).

- (28) *dâšt az ânjâ rad mi-šod*  
 PROG.PST(3SG) from there passing IMPV-become.PST(3SG)  
 'He was passing by there' (Adibifar 2016, g2\_f\_02)
- (29) *faqat dâšt-am komak mi-kard-am*  
 just PROG.PST-1SG help IMPV-DO.PST-1SG  
 'I was just helping' (Davari and Kohan 2017, glosses modified)

<sup>9</sup> I am grateful to Shirin Adibifar for native-speaker intuitions on these examples. My analysis differs from many others in that I do not ascribe any aspectual value to *mi-* when it occurs with a present stem.

<sup>10</sup> Presumably for these two reasons, much of the relevant literature on Persian does assume an aspectual contribution of *mi-* in the present tense, and glosses it as e.g. durative (Taleghani 2008), or imperfective; see Davari and Kohan (2017) for discussion.



- (30) *dâr-i*      *či*      *kâr*      *mi-kon-i?*  
 PROG.PRS-2SG what work IND-do.PRS-2SG?  
 'What are you doing?'  
 (31) *dah sâl=e*      *dâr-e*      *piano mi-zan-e*  
 ten year=COP.3SG PROG.PRS-3SG piano IND-hit.PRS-3SG  
 'He has been playing piano for ten years'

The last example makes it clear that this is not merely objective marking of 'continuous/progressive aspect', but involves more subtle aspects of speaker's stance towards the assertion being made. Obviously the subject in (31) has not been playing the piano continuously (without interruption) for ten years. Rather, the speaker has chosen to portray an activity undertaken at regular intervals in terms of a continuous process.

Davari and Kohan (2017) point to further semantic nuances expressed by progressive *dâştan*, which are not immediately derivable from a purely aspectual sense. With inherently non-progressive verbs, such as 'fall', or 'die', the use of the *dâştan*-auxiliary expresses prospective aspect 'is about to' (examples from Davari and Kohan 2017, glosses adapted).

- (32) *dâr-e*      *mi-mir-e*  
 PROG.PRS-3SG IND-die.PRS-3SG  
 'He is about to die'  
 (33) *begir=eš*      *dâr-e*      *mi-oft-e*  
 hold.PRS.IMP=3SG.P PROG.PRS-3SG IND-fall.PRS-3SG  
 'Hold it! It is about to fall'

In the past, the *dâştan*-progressive may express a prospective state as the temporal framework within which a punctual action occurs:

- (34) *dâšt-am*      *mi-raft-am*      *ke u zang zad*  
 PROG.PST-1SG IMPV-go.PST-1SG CPL 3SG ring strike.PST.3SG  
 'I was about to go when he called' (Shirin Adibifar, p.c.)

Davari and Kohan (2017) interpret these non-progressive uses of the *dâştan* progressive as evidence of increasing 'subjectification' of the aspectual marker, drawing on Narrog's (2012) account of an increase in 'speaker orientation'. The origin of the *dâştan* progressive are obscure. Jügel (2015: 155) finds no evidence for early grammaticalization of 'have/hold' as an auxiliary in his Middle Persian corpus. Davari and Kohan (2017) cite Dehghan (1972), who mentions written examples from the beginning of the 20th century as the earliest attestations. It is of course likely that these constructions had been available in the spoken language for centuries, but have only relatively recently been committed to writing.

Davari and Kohan (2017) see the point of origin for this construction in the reanalysis of a construction involving a centre-embedded relative clause containing the verb *dâştan*, which is reanalysed as a single clause, though this remains to be confirmed in a more comprehensive survey. As an example for this kind of bridging context, they cite among others the following:

- (35) *parde=i*      *ke*      *dâšt-and*      *bâz*      *mi-kard-and*  
 curtain=INDEF that have.PST-3PL open IMPV-do.PST-3PL  
 'The curtain that they had, they were opening' > 'They were opening the curtain'

It should be noted that the *dâştan* progressive is not fully integrated into the TAM system of Persian, because it lacks a negated form.

A functionally similar renewal of aspect has also occurred in the Behdinî dialect of Northern Kurdish (the dialect spoken in the northernmost parts of Iraq bordering on the Syrian and Turkish borders). In Behdinî, an additional aspectual distinction is available in the present tenses to indicate ongoing and immediate activity. Thus (36) contains the particle *yê*, and indicates immediate and ongoing activity,<sup>11</sup> in contrast with (37), lacking the particle, which stresses habitual, rather than current, activity:

- (36) *Azad yê*      *l=Duhok-ê*      *şul*      *di-ke-t*  
 Azad PROG.M.SG in=Dohuk-OBL.F work IND-do.PRS-3SG  
 'Azad is working (now) in Duhok'  
 (37) *Azad l=Duhok-ê*      *şul*      *di-ke-t*  
 Azad in=Dohuk-OBL.F work IND-do.PRS-3SG  
 'Azad works (regularly) in Duhok'

The etymological source of the progressive particle in (36) is quite different from the source of the Persian *dâştan* progressive (unlike Persian, Northern Kurdish entirely lacks a lexical HAVE-verb). The particle *yê* originates in a linking particle (known as the *ezafe* in Iranian linguistics), used among other things to link adnominal attributes to nouns. The main evidence for associating the aspectual particle of (36) with the *ezafe* is that both inflect for number and gender in a very similar manner. Thus if the subject in (36) were female, the particle would have the form *ya*, and if the subject were plural, *yêt*. Originally the *ezafe* was of pronominal or demonstrative origin, and this appears to be the source of its function here; presumably it grew from some kind of cleft construction ('Azad—the one working in Dohuk') though this remains speculative (see Haig 2011 for discussion). Whatever the source, the result is that in the present tense an additional aspectual distinction has been added to the verb system, achieving a similar result to the *dâştan* progressive in Persian. The parallels to the Persian developments also extend to the constraint against negation: the innovated progressive of Behdinî is likewise ungrammatical with a negated main verb.

#### 4.4 SUMMARY

This chapter has surveyed some of the grammaticalization processes that can be traced back across two and a half millennia of Iranian languages, focusing on the contrast between the grammaticalization of person indexing for subjects from

<sup>11</sup> It is notable that in colloquial Behdinî, the 'progressive' form is exceedingly frequent, probably more so than the simple form in (37). Its claimed aspectual force is thus weak, and only really emerges when speakers wish to make a contrast. However, the subtle nuances of usage remain largely unresearched.

erstwhile clitic pronouns and the non-grammaticalization of the cognate set of pronouns in object function. Despite identical input material, the trajectories of these two processes have been very different. I argue that in the case of object clitics, for well over two thousand years the object clitic pronouns have remained just that: clitic pronouns, with little further development towards object agreement. From their earliest Old Iranian origins the clitic object pronouns were prosodically deficient, bound elements. In the course of their subsequent development they have undergone changes in placement principles, and formally may superficially resemble affixes, but there has been no evident shift towards becoming obligatory indexes (agreement) rather than alternating indexes. The subject pronouns, on the other hand, have demonstrably developed into agreement markers in some languages. Although we lack evidence for the precise pathway of development, I have suggested that the process was inaugurated through an increase in frequency of these pronouns already in Middle Iranian, in which the clitic pronouns began to occur in contexts where previously pronouns would have been dispreferred (in same-subject coordination, for example).

As mentioned, much of the inherited inflectional morphology (case, gender, and TAM categories) eroded in a relatively short transition between the Old and Middle Iranian periods, and the history of grammaticalization can to some extent be seen as the gradual reacquisition of lost morphological categories, for example case, modality, and aspect (though the restitution of gender appears to be rare). Notably, across different Iranian languages, these processes do not necessarily draw on cognate source material, and may in fact be absent altogether. Thus some languages lack grammaticalized aspect in the present tenses; others have not restituted structural case marking (e.g. still lack an accusative or genitive case). Where these inflectional categories have been reinstated, the process has been slow, and most examples still bear obvious traces of the source construction, e.g. through stress patterns (Persian = *râ*), or traces of syntactic mobility of the marker (e.g. subject agreement in Central Kurdish), or a failure to cover all the relevant slots in a paradigm (e.g. the failure of Persian *dâštan*-progressive to occur with negated main verbs). There are thus quite discernible structural differences between the 'inherited' morphology (where it has survived at all) and the 'innovated' inflectional morphology of Iranian (Haig 2008: 96; to appear, b). To reach that state, then, the inherited morphology itself must have gelled over an exceedingly long developmental period—far longer than the two thousand years of attested history of Iranian languages available for our perusal. It is thus scarcely surprising that the origins of (for example) inherited verbal person agreement morphology lie far beyond the limits of historical records, and are likely to remain unknown to historical linguists. Inflectionalization is evidently a process that requires millennia, not centuries, to achieve, though paradoxically, its loss can be quite rapid, even catastrophic.

## 5

## Grammaticalization in the languages of Europe

ÖSTEN DAHL

## 5.1 INTRODUCTION

Is it possible to give a characterization of grammaticalization processes in European languages? To make any sense, such a characterization must tell us not only what grammaticalization in Europe is like, but also how it differs from grammaticalization in other parts of the world.

The problem that arises is not primarily that we know too little about grammaticalization in Europe; it is rather that we know so little about it elsewhere. For many European languages, we can follow their history back for more than two millennia, with abundant written documentation. Outside of Europe, this is the case only for a very limited number of languages. It is also the case that the scientific study of language has been very much focused on the major languages of Europe, and most linguists have been native speakers of European languages. The study of grammaticalization has been no exception, with the standard examples tending to be taken from the history of 'Standard Average European' (SAE) languages. So it is unlikely that we will find anything which deviates radically from conventional wisdom by just trying to see what is found in European languages.

Excluding chance as an explanation, similarities in developments between languages could be explained either (i) through similarities in preconditions—either internal, i.e. shared structural properties, or external—shared ecologies, or universal cognitive properties, or (ii) through influence due to language contact. The first type of explanation, which is generally favoured in typology, allows for generalizations over a set of languages—but it demands that we treat the members of the set as independent cases, in order to exclude influence through contact. By contrast, in areal linguistics, scholars normally want to exclude parallel independent developments. There are stumbling blocks of different kinds here. Someone who is looking for