

The pronoun-to-agreement cycle in Iranian: subjects do, objects don't.

Geoffrey Haig

Pre-print manuscript, under review for:

Larson, Richard & Sedigheh Moradi & Vida Samiian (eds.) *Advances in Iranian linguistics*. Amsterdam: Benjamins.

Comments welcome

1 Introduction

The view that long-term grammatical change is cyclic in nature was widespread among linguists of the late nineteenth and early twentieth centuries, for example Georg von der Gabelentz, Edward Sapir, and Otto Jespersen. More recently, the idea has resurfaced in derivational approaches to diachronic syntax, where for example Jespersen's Cycle in the emergence of negation markers has been re-framed in Minimalist terms (van Kemenade 2000, van Gelderen 2011b, among many others). Perhaps the most comprehensive study in this direction is van Gelderen (2011b), who analyses a number of different kinds of diachronic change in terms of internally-motivated cycles. In this chapter, I focus on what van Gelderen (2011b) refers to as the "head-marking cycle". This cycle begins with a pronoun, an element filling an argument position in syntax, which develops into an agreement marker, hence lacking a theta role, before finally eroding to zero. The cycle then begins afresh, with a new element emerging in the pronoun function. According to van Gelderen (2011a, b), this cycle is attested both for subject pronouns, and object pronouns, and can be attributed to universally operative and internally motivated principles within the Minimalist framework.

The Iranian languages, with some 2500 years of attested history, and dozens of surviving modern languages, provide a generous window for observing the kinds of long-term diachronic changes which cyclic approaches

presuppose. In this chapter I will briefly sketch what is known regarding the "agreement cycle" in West Iranian languages, and evaluate the cyclic model of van Gelderen (2011b) against the Iranian evidence. I consider the development of subject pronoun to subject agreement marker, and for object pronouns to agreement marker respectively. The most striking finding is the almost complete absence of such a development for object pronouns, all the more surprising given the fact that the assumed preconditions for the grammaticalization of object agreement, namely cliticization of the relevant pronouns, has been available for millennia. But to the best of my knowledge, the early cliticization of object pronouns has not yielded object agreement anywhere in Iranian. For subject pronouns, on the other hand, uncontroversial cases of agreement markers developing from erstwhile clitic pronouns are attested, though only a small section of the assumed cycle is actually historically verified.

The asymmetry in the outcomes of subject and object pronoun grammaticalization in Iranian is not an isolated phenomenon, but reflects a widespread typological tendency. Nevertheless, much of the relevant literature continues to assume a unified grammaticalization pathway for subject and object pronouns. Thus Schiering (2005: 45) simply states that "cliticized subject pronouns can become agreement affixes cross-referencing the subject NP; cliticized object pronouns can become agreement affixes cross-referencing the object NP", Siewierska (2004) suggests that the grammaticalization of pronouns towards agreement is "a continuous process on-going in all languages in all times", without differentiating object and subject pronouns, while van Gelderen (2011b), assumes the existence of two cycles, a subject cycle and an object cycle, but provides no explanation for the evident cross-linguistic differences. In this chapter, I will briefly outline the essence of van Gelderen's (2011b) proposals, then present a summary overview of the relevant data from Iranian. Finally, I will address the adequacy of the Minimalist cyclical approach to the grammaticalization of agreement. Given the scope of the issues involved, I can do little more than

sketch the main arguments and present what I trust is a reasonably representative cross-section of data. Some of the issues here have been dealt with from a cross-linguistic perspective in Haig (2018a) and Haig (2018b), to which the reader is referred for additional arguments and data.

2 The agreement cycle according to van Gelderen (2011b)

The idea that verbal agreement markers arise from originally free pronouns was popularized by Givón (1976), and is rooted in the observation that in many languages, paradigms of agreement affixes often demonstrate close phonological similarities to the corresponding sets of personal pronouns (Siewierska 2004: 251-254, Haig 2018b). A natural explanation for these similarities is that the agreement affixes represent the grammaticalized remnants of erstwhile free pronouns. Indeed, this assumption is widely regarded as a given. As Siewierska (2004: 251) notes, "everyone acknowledges that person clitics and affixes typically evolve from independent person markers [free pronouns, GH]".

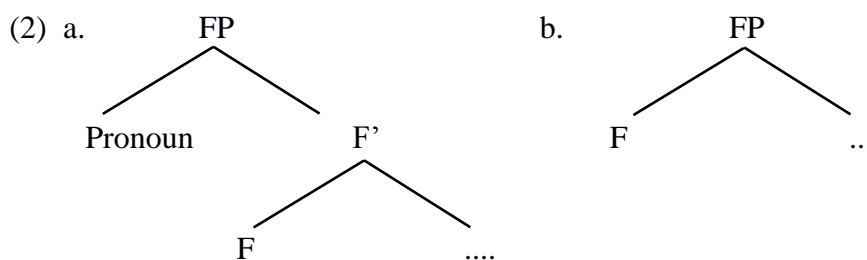
Van Gelderen (2011b) likewise assumes that pronouns are a common diachronic source for agreement morphology. In her framework, the development is seen as one of several cyclic processes in the creation of inflection, behind which quite abstract and very general principles can be identified. The theory is powerful in the sense that superficially distinct processes are considered as manifestations of a small number of very general principles. The most relevant principles in the present context are so-called Principles of Economy, which are operative in the resolution of "ambiguous structures" that arise in the derivation of syntax (van Gelderen 2011b: 13). Despite the name, Principles of Economy are not general cognitive principles geared to optimizing processing costs. Rather, they are principles specific to "I-Language", rather than performance-based principles relevant to "E-Language". In what follows, I will only consider two Principles of Economy,

The Head Preference Principle (HPP) and Feature Economy (FE). The Head Preference Principle (HPP) is given in (1):

(1) The Head Preference Principle (HPP)

Be a head, rather than a phrase (van Gelderen 2011b: 13)

More generally, "whenever possible, a word is seen as a head rather than a phrase" (van Gelderen 2011b: 13). The effects of the HPP can be schematically illustrated in (2), where FP stands for any functional category (here illustrated with a pronoun). When a functional element such as a pronoun or an adposition is merged, the HPP will yield an interpretation (2b), rather than (2a), if a speaker is exposed to evidence compatible with either (van Gelderen 2011b: 13).



(van Gelderen 2011b: 13, ex. 16)

In terms of syntactic derivation, the HPP translates into a preference for head, rather than specifier position. With regard to the difficulties of distinguishing specifiers from heads, van Gelderen (2011b: 14) provides the following criteria: "Specifiers are full phrases and can be modified and coordinated, and they occur in certain positions; a coordinated or modified element is never a head, and head movement is usually recognizable."

The HPP is relevant to a number of historical changes, for example demonstrative *that* > complementizer *that*, adverb > aspect marker, or pronoun > agreement. Of course changes of this kind are regularly cited in

the grammaticalization literature, and accounted for in terms of a cline from ‘lexical to grammatical’, or ‘less grammatical to more grammatical’. The Minimalist account of van Gelderen (2011b) is an attempt to integrate these observations into a more formalized framework, and define more rigorously the somewhat vague notion of ‘more grammatical’.

The second kind of principle that is relevant for the subject agreement cycle concerns the nature of the features associated with the merging elements. Pronouns typically involve features that are both relevant for the semantic interpretation of an utterance, but also for the correct spell-out of associated inflectional morphology. In the version of Minimalism espoused in van Gelderen (2011b: 17), features are considered to be either ‘interpretable’ or ‘uninterpretable’:

Starting with Chomsky (1995), the features relevant for and accessible during the derivation are formal. Formal features can be interpretable (relevant to the semantic interface) or uninterpretable (only relevant to move elements to certain positions). Interpretable features are acquired before uninterpretable ones [reference omitted, GH¹], but are later reinterpreted as uninterpretable, triggering the functional/grammatical system. The same happens in language change.

Uninterpretable features are preferred because they provide the impetus for the derivation: "If you select two words from the lexicon with only

¹ The reference deleted from this citation is to an internet publication, but the source provided in van Gelderen (2011b) is no longer identifiable online, hence I have removed it; it is not relevant to the arguments at hand.

interpretable features, they will not interact or merge." (van Gelderen 2011b: 20)²

With regard to the pronoun-to-agreement shift, the changes can be schematically illustrated as follows, where ‘phi’ abbreviates the person values first, second and third person (1,2,3), ‘i’ abbreviates ‘interpretable’, and ‘u’ is ‘uninterpretable’:

(3) emphatic full pronoun head pronoun agreement
 [i-phi] > [i-phi] > [u-1/2], [i-3] > [u-phi]

The cline sketched in (3) is driven by a Feature Economy Principle, formulated in van Gelderen (2011b: 17) as "Minimize the semantic and interpretable features in the derivation". Van Gelderen (2011b) also distinguishes between a feature ‘first/second person’ and ‘third person’, a move motivated by the fact that pronouns with these features grammaticalize at different rates towards agreement; in general, first and second person pronouns spearhead the development, becoming uninterpretable earlier, while third person pronouns apparently lag behind. On her view, the pronouns of the first and second person entail "pure phi-features (person and number" (van Gelderen 2011b: 74). Third person pronouns on the other hand, encode additional features, though the nature and number of these features is a matter of typological variation. Typically they involve gender, and deixis; the latter would be particularly true of languages lacking dedicated third person pronouns, instead relying on forms identical to distal demonstratives. The forms with the simplest feature specification are therefore first and second person forms, and these are the forms which are thus more likely to shift their features from interpretable to uninterpretable.

² The Feature Economy Principle outlined here is actually considered an offshoot of ‘Late Merge’, but I have omitted the relevant argumentation here, see van Gelderen (2011b: 14-17) for details.

Taken together, the Head Preference Principle and Feature Economy conspire to nudge free pronouns, as phrase-projecting carriers of interpretable features, to become exponents of non-projecting heads with uninterpretable, but syntactically relevant, features: agreement morphology. Of course in order to complete the cycle, the agreement morphology must further develop to zero, before being replaced by innovated material realizing the relevant features, thus yielding a complete cycle. The shift from agreement head to zero may apparently be triggered by the ‘stacking up’ of additional material in the same slot, as additional functional heads accrue in the same position leading to opaquely fused morphology, and ultimately complete loss of the original material (van Gelderen 2011b: 19-21). However, this stage of the cycle (the loss of agreement) will not be considered further here.

It should be emphasized that the above outline ignores much of the technical details of van Gelderen’s proposals, which would go beyond what can reasonably be accommodated here.³ The crucial point is that the developments are formulated in terms of purely syntactic and very general principles, which should in principle be universally operative.

2.1 The agreement cycle with subject pronouns.

Strictly speaking, the starting point of the agreement cycle in (3) are ‘emphatic pronouns’, but I will assume here ‘full pronouns’ for reasons of brevity. Full pronouns exhibit certain properties, which are exemplified with pronouns from Hindi in (4). They can be modified (4a,b), or carry a special focus particle (4c), they inflect for case in the same manner as other DPs, and show similar positional distribution to DPs, and can also be coordinated. These are typical diagnostics for the ‘nominal’ nature of such pronouns: "In

³ I also ignore the discussion of so-called polysynthetic languages, for which somewhat different mechanisms are required (van Gelderen (2011b: 43-44), see Corbett (2006: 100-113) for critical discussion of ‘pronominal affixes’, which is relevant to the notion of polysynthesis.)

terms of features, the pronouns and DPs are full phrases at this stage and carry the traditional Case and phi features." (van Gelderen 2011b: 47)

- (4) *mē* 'I', *tum* 'thou', *woo* 'she/he', *ham* 'we', *aap* 'you', *woo* 'they'
- a. *ham log* 'we people'
 - b. *aap log* 'you people'
 - c. *mē hii* [1SG-FOC] 'I'

[Hindi, van Gelderen 2011b: 45]

The transition to the second stage, so-called 'head pronouns', is not abrupt; instead we find various intermediate phases. The so-called 'subject pronouns' of English (*I, you, he, she* etc.) differ, for example, from the so-called 'object pronouns' *me, you, him, her* etc. with respect to some of the relevant parameters. The subject pronouns are generally unstressed, have less syntactic freedom (they are restricted to a position immediately preceding a finite verb, separable from it only by a small set of adverbs), not available in isolation or in focus constructions such as *as for X, ...*, and are dispreferred in coordination (a combination such as *?they and we*, for example would be avoided in my dialect of spoken English). On van Gelderen's (2011b) approach, this is evidence of an initial move down the Subject agreement cycle, from full to head pronoun. More advanced developments are found in colloquial French, where the weak series of pronouns *je* etc. frequently double an overt subject NP, as in (5) and (6):

- (5) *une omelette elle est comme ça*
 an omelet she is like this

'An omelet is like this.' [Spoken Swiss French,

Fonseca-Greber 2000: 335, cited in van Gelderen 2011b: 52]

- (6) *Moi je suis un blogueur*
 Me I am a blogger

‘I am a blogger.’ [Colloquial French, van Gelderen 2011b: 53]

There is evidently good reason to consider the weak pronoun set of French *je* etc. to be functionally distinct from the free pronouns of, e.g. Hindi discussed in (4). In van Gelderen’s terms, they can be considered a further step on the cycle towards agreement.

The final stage of fully obligatory agreement marking is illustrated by languages such as German or Latin, exhibiting obligatory agreement with subjects, regardless of any pragmatic considerations. It is worth pointing out that precisely this kind of canonical agreement (Corbett 2006) often cannot be directly traced to a pronominal predecessor. Van Gelderen (2011b) does not actually provide a convincing example of the entire cycle, but instead takes the structures from (often unrelated) languages as representatives of the various stages of the assumed cycle. The assumption seems to be that the weak pronouns of French illustrated in (5-6) will somehow eventually morph into obligatory agreement affixes, given sufficient time, but clear evidence of such a process is hard to find, as Siewierska (1999) had already noted. As it turns out, evidence for the final stages of this process can be found in Iranian (see Section 3.1).

2.2 *The agreement cycle with object pronouns.*

Direct evidence for the object agreement cycle is hard to come by. Van Gelderen (2011b) illustrates it with the following fictitious example:

- (7) a. *I saw yesterday her (and him)*
 b. *I saw’r yesterday (*and him)*
 c. *I saw (’r) **HER**.*

[Fictitious English, van Gelderen 2011b: 88]

In (7a), the pronoun is syntactically and prosodically independent (separable from the verb by an adverb), bears a theta role, and can be coordinated. In (7b) it has lost positional freedom and prosodic independence, and also the ability to coordinate. In (7c) we observe the possibility of doubling the attached pronoun through an additional "emphatic pronoun". At this stage the attached pronoun may erode to zero, "and the cycle can start over again." (van Gelderen 2011b: 88). The stages of these developments are sketched in the form of "a possible cline" in (8):

(8) phrase > head > agreement > zero
 [i-phi] [i-phi] [u-phi]
 [u-Case]

(van Gelderen 2011b: 88)

Van Gelderen (2011b: 90) points to the considerable cross-linguistic diversity in object agreement systems: "With respect to object agreement, there is enormous diversity as to what starts the cycle. Animate and definite object pronouns of all persons are reanalyzed as object agreement but there is no obvious pattern." Similarly, the initial structural configuration for objects is less clear-cut, because current conceptualizations of VP structure yield somewhat different analyses (cf. the alternative options in van Gelderen 2011b: 89-90). But in essence, both the HPP and Feature Economy are considered active in driving the developments, just as they are with subject pronouns. A problem nevertheless arises with regard to Feature Economy, because object agreement is often sensitive to animacy, and definiteness, features that are related to person, but not identical to it. Van Gelderen (2011b: 90) assumes that definiteness is related to the presence of uninterpretable [ASP] (aspect) features on the verbal head governing the object, though I find the connection somewhat tenuous. But apart from the apparent difficulties in identifying the favoured starting configuration for the cycle, van Gelderen (2011b) assumes that the object agreement cycle can be

motivated in a manner that parallels that of the subject agreement cycle, and there is thus no expectation that the outcomes of the two cycles will be any different.

3 *Clitic pronouns and agreement in Iranian*

Subject agreement via affixes on the verb is present in most, perhaps all, modern Iranian languages, though it may be absent for past transitive clauses. For ease of exposition, we may take modern standard Persian, where the paradigm of subject agreement suffixes is provided for the present indicative in Table 1.

Table 1: Subject agreement suffixes in Persian (present indicative of *xordan* ‘eat’)

	SINGULAR	PLURAL
1	<i>mi-xor-am</i>	<i>mi-xor-im</i>
2	<i>mi-xor-i</i>	<i>mi-xor-id</i>
3	<i>mi-xor-ad</i>	<i>mi-xor-and</i>

The markers themselves are suffixal, rather than clitics: they are restricted to a specific slot (immediately following the verb stem), they are obligatory in the sense that they are required by a particular syntactic configuration, irrespective of the presence or absence of a full NP subject in the clause, and according to Kahnemuyipour (2003: 374-375), are ‘cohering’ suffixes, i.e. part of the phonological word. In the sense of van Gelderen (2011b), they are syntactic heads, associated with uninterpretable phi-features.

The origins of these suffixes are obscure. Persian has exhibited some form of agreement suffixes in comparable environments for as long as we have attested records. Thus if they are the endpoint of a grammaticalization process

that began with a free pronoun, the earlier phase of the development lie beyond the bounds of what can realistically be reconstructed.

Along with the suffixes of the type illustrated in Table 1, Persian and the majority of other Western Iranian languages exhibits a second set of prosodically dependent person and number marking morphemes, often referred to as clitic pronouns. Although the paradigms are not fully identical, they can reasonably be considered cognate with the pronominal clitics attested in Middle West Iranian language such as Parthian, Middle Persian and Bactrian (Jügel 2015). The Middle West Iranian clitics are provided, together with a selection of contemporary West Iranian languages in Table 2.⁴

Table 2: Clitic pronouns in Western Iranian languages

	Appr. 2000 years BP	Contemporary West Iranian languages				
	Middle West Iranian	Persian	Vafsi	Hawrami	Sivand	C. Kurd. Sanandaj
1SG	= <i>m</i>	= <i>am</i>	= <i>om</i>	=(<i>i</i>) <i>m</i>	= <i>em</i>	= <i>im</i>
2SG	= <i>t</i>	= <i>at</i>	= <i>i</i>	=(<i>i</i>) <i>t</i>	= <i>et</i>	= <i>it</i> / = <i>o</i>
3SG	= <i>š</i>	= <i>aš</i>	= <i>es</i>	=(<i>i</i>) <i>š</i>	= <i>eš</i>	= <i>ī</i>
1PL	= <i>mān</i>	= <i>mān</i>	= <i>oan</i>	= <i>mā</i>	= <i>emā</i>	= <i>mān</i>
2PL	= <i>tān</i>	= <i>tān</i>	= <i>ian</i>	= <i>tā</i>	= <i>etā</i>	= <i>tān</i>
3PL	= <i>šān</i>	= <i>šān</i>	= <i>esan</i>	= <i>šā</i>	= <i>ešā</i>	= <i>yān</i>

This paradigm of clitic pronouns has proved remarkably robust, surviving across at least 2000 years in recognizable form in the majority of West Iranian languages, though lost in Zazaki, Northern Kurdish, Gilaki and Mazanderani.

⁴ Sources for the languages other than Persian: Middle West Iranian: Jügel (2015:222), see also Korn (2009) for historical details on this paradigm; Vafsi: Stilo (2018: 695, Table 5E); Hawrami: MacKenzie (1966: 25); Sivand dialect: Lecoq (1979: 40); Central Kurdish of Sanandaj: Öpengin & Mohammadirad (to appear). 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.

Note that these clitic pronouns are not simply phonologically reduced forms of today's full pronouns. Rather, they are the reflexes of a distinct set of clitic non-nominative pronouns, of which the corresponding full pronouns have disappeared. From a synchronic perspective, they are not relatable to the full forms of the pronouns via predictable phonological rule in any of the languages listed.

For reasons outside the purview of this chapter, in the Old Iranian period, the clitic pronouns in Table 2 came to be used as **subject** pronouns with past transitive verbs. I follow a long tradition in referring to this function as 'A', meaning 'transitive subject', but it should be borne in mind that the only transitive subjects that occurred with this kind of pronoun were those associated with verbs built on the old participial stem, generally referred to as the 'past stem'.⁵ In the next section I briefly sketch the workings of these subject clitic pronouns, while in 3.2, I discuss the same set of pronouns in object function.

3.1 *Clitic pronouns indexing subjects (A)*

In Old Iranian, and well into Middle Iranian, the subject clitic pronouns were in complementary distribution with a co-referent NP subject. Example (9) from Middle Iranian illustrates a clitic pronoun A, while (10) has a NP in the A role, and no clitic pronoun:

- (9) *čē=t* *ātaxš* *ī* *man* *pus* *ōzad*
 because=2SG:A fire of my son extinguish.PST.3SG
 'because **you** extinguished the fire of my son [...]'

⁵ In fact they also serve as 'subject' agreement for a number of other constructions across the West Iranian languages, including predicates of experience, desire, possession, and physical and mental states such as hunger. In these functions, the use of the clitic pronoun as a subject index is independent of tense, see Haig and Adibifar (in press).

[Middle Persian, Haig 2008: 124]

- (10) *pas ošbām oy az pidar bōxt [...]*
 then **ošbām:A** 3SG from father rescue.PST.3SG
 then **Ošbām** rescued her from (her) father [...]
 [Zoroastrian Middle Persian, Jügel 2015: 410, glosses added]

Jügel (2015: 400) notes the general lack of clitic doubling in Middle Iranian,⁶ underscoring the pronominal nature of the clitics at this stage. Another important indicator of their pronominal nature is that they could be omitted in contexts where the identity of the subject is pragmatically recoverable, for example in same-subject clause chaining. Example (11) has an overt clitic pronoun for the A of the first clause, and zero for the co-referential A of the subsequent clause:

- (11) a. *u=š ardawān ožad [...]*
 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_i killed Ardawān [...] and (he_i) took his daughter as wife’
 [Zoroastrian Middle Persian, Jügel 2015: 411, glosses added]

The available evidence thus supports a pronominal interpretation of the clitic pronouns, because (i) they cannot co-occur with a co-referent NP in the same clause, and (ii) they may be omitted in precisely those environments that free

⁶ Jügel (2015: 396-399) notes a small number of Middle Persian examples where the A-clitic is doubled by an overt A in the clause (94 attested in a corpus 6815 clauses). Some may be attributed to scribal errors or other problems of interpretation and transmission.

pronouns would likewise generally be omitted (e.g. corereferential deletion in coordinate clauses).

However, there are criteria for distinguishing among different kinds of pronoun, rather than assuming a general binary split between an agreement marker on one hand, and a pronoun on the other (cf. Jügel and Samvelian, this volume, for the latter view). Van Gelderen (2011b) recognizes a distinction between free pronouns, and "head pronouns". The former have the same word order freedom as lexical NP's, are stressable, focusable, and can be modified and coordinated (cf. discussion in connection with (4) above). Head pronouns, on the other hand, lack at least some of these features. The clitic A-pronouns of Middle Iranian would most likely qualify as "head pronouns": their position is fixed through the second-position principle governing clitic placement in Middle Iranian, and it seems unlikely that they were stress-bearing, or capable of expressing contrastive focus.

There is a further criterion for distinguishing between free pronouns and head pronouns, discussed in Haig (2018a: 67). As mentioned, free pronouns are characteristically omitted under conditions of pragmatic identifiability of the referent, and this can be considered a general feature of pronouns, though famously subject to cross-linguistic variability (see Torres Cacoullos and Travis, in press, and Haig and Adibifar, in press). Thus full pronouns are typically characterized by a pragmatically-determined alternation with zero. The clitic subject pronouns of Middle Iranian could also be omitted, for example in the second conjunct of same-subject clause sequences (cf. (11b) above). But in fact, several Middle Iranian examples illustrate clitic pronouns in contexts where pronouns would not normally be expected, for example the following:

- (12) *ēk, ke=š man 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, glosses added]

The subject pronoun =š attaches to the relativizier *ke*, although resumptive pronouns are generally not required in Iranian subject relativization. Jügel and Samvelian (this volume, §2.1) also note the propensity for clitic pronouns to occur in same-subject sequences of main and embedded clause in their Middle Persian corpus, again precisely an environment where zero would be the expected option. Jügel and Samvelian (this volume, Fig. 5) provide figures from the analysis of a single Middle Persian text. The numbers of zero subjects in past transitive clauses is significantly lower than in present transitive clauses (44% versus 72%). A Fisher’s exact test of this difference yields a value of 0.0001, highly significant when compared with a significance value of $p < 0.05$.⁷ What I would provisionally conclude from these findings is that the past transitive clauses avoid zero expression of subjects to a greater degree than the present transitive clauses.

The overall figures for the Middle Persian corpus investigated in Jügel (2015: 326, Table 5.4) indicate that around 44% of all past transitive clauses contained a clitic pronoun exponent of the subject (N=6815). Comparing this figure with the percentage of overt pronouns in transitive clauses of other languages which allow referential null subjects is revealing: In contemporary spoken Persian (Adibifar 2016), overt A pronouns occur in 8% (N=603) of

⁷ My calculations are based on a comparison of non-zero realizations (NP, pronouns) versus zero-realizations of A-arguments in two conditions, past versus present tense (based on the figures in Fig. 5, Jügel and Samvelian, this volume). Interestingly, with intransitive subjects there is an inverse effect of tense, with zero realizations being more frequent in the past than in the present. This appears to be linked to the more narrative nature of the past-tense sections of the text (cf. Jügel and Samvelian, this volume), which would favour topic continuity over longer stretches, hence zero expression. If this is the case, then it further heightens the significance of the reduced levels of zero realizations for transitive subjects in past tenses.

the transitive clauses, in Cypriot Greek (Hadjidas & Vollmer 2015) we find just 4% (N=494), and in Northern Kurdish (Haig & Thiele 2015) 29% (N=422).⁸ For these languages, and indeed most others that allow null referential subjects, the favoured form of expression for transitive subjects is zero, not pronominal. The Middle Persian figure of 44% (a conservative estimate) is thus significant, and suggests that these so-called ‘pronouns’ were of a qualitatively different kind to the free pronouns.

Jügel and Samvelian (this volume) also note the difference, and assume that it is due to the lack of subject agreement morphology on past transitive verbs. This suggests that the triggering factor for the grammaticalization of clitic pronoun subjects towards agreement markers was essentially structural: the loss of an old paradigm of suffixal agreement morphology is compensated by recruiting a new paradigm from the available clitic pronouns. This scenario is in line with Fuß’ (2005) claims regarding the motivation for the emergence of subject agreement as compensation for defective agreement paradigms. In principle I find this plausible, and the frequency data from Middle Iranian provide empirical support for such a view. To what extent additional explanations in terms of re-analysis of ‘hanging topic’ constructions are required remains an open question (see Schnell 2018, among many others, for critique of the ‘dislocated topic’ approach to the emergence of agreement, and Jügel and Samvelian, this volume, for an attempt to justify it for Iranian).

The system of indexing the A through a pronominal clitic has disappeared in some contemporary Iranian languages, notably Persian, but in others it has survived remarkably well. However, in some languages the nature of the clitic pronoun has changed. In Central Kurdish, the pronominal clitic has 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). Along with a functional shift,

⁸ The comparatively high figure for Northern Kurdish 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.

the clitics have changed their position, from the clause-second position of Middle Iranian to a VP-based placement (cf. Haig’s (2008: 336) ‘rightward drift’ of clitic placement in Iranian). This is illustrated with the following examples from the Mukri dialect of Central Kurdish: (13) shows the co-occurrence of a pronominal clitic with a definite NP subject, (14) an indefinite, non-specific subject, and (15) a pronominal subject.

- (13) *qerewol-ān kut=yān*
guard-PL say.PST=3PL
 ‘The guards said ...’ [Öpengin 2013: 307, cited in
 Öpengin & Mohammadirad, to appear]

- (14) *hîç kes řā=y-ne-de-girt-im*
no person PVB=3SG-NEG-IPVF-keep.PST-1SG
 ‘Nobody would let me in (their house).’ [Öpengin 2013: 51,
 cited in Öpengin & Mohammadirad, to appear]

- (15) *min ne=m-dît*
1SG NEG=1SG-see.PST
 ‘I did not see him.’ [Öpengin 2013, cited in
 Öpengin & Mohammadirad, to appear]

There is a broad consensus in the relevant literature that the pronominal clitics in Central Kurdish are exponents of an agreement relation (see Samvelian 2007; Haig 2008, 2018b; Öpengin, Forthc.; and Öpengin & Mohammadirad, to appear; see Dabir-Moghaddam 2003 for examples from other West Iranian languages).

In a number of other West Iranian languages, however, the clitic remains in complementary distribution with a coreferent NP subject. This appears to be the situation in the Surçî dialect of Northern Kurdish, spoken in Iraqi Kurdistan:

- (16) *min* *la_bo* *xo* *rēnjbār-ak* *girt*,
 1SG.OBL for self labourer-INDEF take.PST.3SG
- hinār=im=a* *jot*, *šiwān-ak=iš=im* *girt*
 send=1SG=DRCT plough shepherd-INDEF=ADD=1SG take.PST.3SG
- ‘I hired a labourer, I sent (him) to the plough, (then) I hired
 also a shepherd.’ [MacKenzie 1962: 228, cited in Öpengin &
 Mohammadirad, to appear]

In this dialect then, unlike the Central Kurdish outlined in (13-15), a subject clitic is not obligatory. In fact there are also clauses in MacKenzie’s (1961) data that contain neither a subject clitic, nor an overt subject NP, indicating that the clitic pronouns are still omissible under pragmatically felicitous conditions.⁹

Finally, we can point to those languages where the clitic pronouns are used for subject indexing, but have lost all positional freedom and occur exclusively on the verb stem itself, thus resembling more closely an affix. This is found with third person subjects in the Kakevendi and Aleshtar dialects of Lak, where the subject clitic only occurs on the verb, regardless of the availability of other potential hosts in the clause:

- (17) *tamām* *māhī-la* *hwārd=ē*
 all fish-PL eat.PST=3SG
- ‘He ate all the fish.’ [Lak of Kakevendi, Öpengin &
 Mohammadirad, to appear]

⁹ See for example “... *sē zēř ta kūna karīřā kird*.” ‘... (he) put three (pieces of) gold on the donkey’s backside’ (MacKenzie 1962: 232).

In the Central Plateau dialect of Semnān, the subject clitics (with past transitive verbs only) have entirely lost their syntactic mobility, and are now restricted to occurring on the verb stem (Haig 2018a).

In sum, across Western Iranian we witness the presence of clitic pronouns indexing past transitive subjects. These pronouns were originally special clitics, rather than free pronouns. They were syntactically constrained in their placement possibilities, and thus presumably lacked the ability of free pronouns to express contrastive focus, or to be coordinated. They also differed from free pronouns in having a higher overall frequency of occurrence, though this requires further investigation. However, they remained pronominal in the sense that they were in complementary distribution with co-referent NP subjects. Pronominal clitics exhibiting very similar properties can still be observed in the Surči dialect of Northern Kurdish, and in at least some dialects of Hawrami, though there are complications here involving word-order variation, and case-marking. For Central Kurdish, on the other hand, and perhaps for Semnān dialect, the clitic pronouns are now fully obligatory agreement markers.

The development could thus be interpreted as traversing a sub-section of van Gelderen's subject agreement cycle, namely that of head pronoun to agreement marker, discussed in (3) above and repeated here for convenience:

- (3) emphatic full pronoun head pronoun agreement
 [i-phi] > [i-phi] > [u-1/2], [i-3] > [u-phi]

As mentioned, the status of van Gelderen's 'head pronoun' is somewhat obscure. Likewise as yet we lack evidence for the assumed split of first and second person versus third person pronouns. But the clitic pronouns that we encounter in Middle Iranian (9-12) are not emphatic pronouns, and arguably distinct from full pronouns. Thus over 2000 years, we find in some, but not all, languages, evidence for the assumed development from pronoun towards agreement marker, albeit only involving the final stages of the cline, and only

attested in some of the languages that have the relevant pronouns. The presumed initial stages, i.e. the development of full pronouns to clitic pronouns, lies beyond the bounds of the historical records. Thus at least 2000 years were required for just the final section of the assumed developmental cline to unfold.

3.2 *Clitic pronouns indexing objects*

The use of clitic pronouns for objects is a feature that characterizes Iranian back to its earliest attestation, so we can assume their presence in Iranian languages for at least 2500 years and probably longer. As such, they are of greater antiquity than the clitic pronouns for subjects (previous section), which only emerged in the wake of the shift to ergative alignment, and only in past tenses (Haig 2008, Jügel 2015). But despite the fact that clitic object pronouns have been around longer, I claim that they have not moved significantly closer towards agreement than their earliest attested forerunners.

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

- (18) *čīd=mān* *pāyēd*
 always=**1PL** protect.PRS.3SG
 ‘(It) always protects **us**’

- (19) [...] *u=š* *hamēw* *bōžēnd*
 [...] and=**3SG** always save.PRS.3PL
 ‘(the Gods) always save **him**’

and morphologically integrated into its host. The best-known counter-examples to this trend come from colloquial spoken Persian, where sporadic instances of clitic doubling can be found. Van Gelderen (2011b: 96) cites examples from Lazard (2006 [1957]), which apparently illustrate that object clitic pronouns in Persian are moving towards agreement (cited from the reprint (2006), and rendered in the colloquial style of transcription):

- (21) (*to xodet miduni ... 'you yourself know...'*)
ke man to=ro duss=et dâr-am
 that I 2sg=ACC loving=2SG have.PRS-1SG
 ‘... that I love you’ [Colloquial Persian, Lazard 2006: 100, 176]

The other examples cited involve third person objects, such as the following:

- (22) (*umadan mixan... 'they came wanting...'*)
baba-jun=o be-gir-an=eš
 father-dear=ACC SUBJ-arrest.PRS-3PL=3SG
 ‘... to arrest dear father’ [Colloquial Persian, Lazard 2006: 176]

Both these examples are taken from written works of fiction (Čubak). An example from authentic spoken Persian is the following:

- (23) *yek pesar-i âmad bā dočarxe ke*
 one boy-INDEF come.PST.3SG with bike COMPL

yeki az zanbil-hā=rā gozast=aš ruye dočarxe=aš
 one of basket-PL=ACC put.PST.3SG=3SG onto bike=3SG.POSS
 ‘a boy came with a bike, then put one of those baskets onto his bike’
 [Adibifar 2016, G2_f_7, 007, cited in Haig 2018b]

Van Gelderen (2011b: 96) considers examples such as (21)-(23) as evidence for the "reanalysis of the verbal object clitic as third person agreement." However, it is not the case that all definite third person objects are accompanied by the corresponding clitic pronoun. The figures from the corpus of spoken Persian in Adibifar (2016) indicate that of the total number of 628 direct objects, just 46 have clitic object pronouns. Among those 46 cases of clitic object pronouns, a sole example, (23) above, involves clitic doubling (Haig 2018b). Thus more than 90% of direct objects are not indexed by a clitic pronoun at all, and doubling of the clitic pronoun with an overt NP is very unusual, at least in this corpus.

Although some examples are acceptable to native speakers, the frequency of direct objects that are accompanied by clitic doubling in actual usage is low. Furthermore, there are quite strict constraints on doubling; Rasekh (2014) notes that doubling the object clitic is not possible with objects that are indefinite, or in focus. My impression is that it is most natural with third person objects, less so with second person objects, and almost unacceptable with first person objects,¹¹ though this requires much more detailed investigation. Van Gelderen (2011b), however, interprets the isolated instances of clitic doubling cited in the literature as evidence for an ongoing shift towards object agreement in Persian. There are several problems with this claim. First, there is actually no evidence that object doubling in contemporary colloquial Persian is an innovation. We simply do not know very much about the colloquial spoken Persian of earlier centuries; it was not written down or recorded. It is quite possible that object doubling has been available as a marginal stylistic device, linked to some specific pragmatic contexts, for centuries, perhaps millennia. Second, even if it should be a comparatively recent innovation, it will not necessarily inevitably proceed

¹¹ I am very grateful to Mohammad Rasekh-Mahand for sharing his intuitions as a native speaker and linguist on these constructions. Of course he bears no responsibility for how I have interpreted them.

towards obligatory agreement. It has been shown that cross-linguistically, some kind of pragmatically conditioned object indexing, as opposed to obligatory agreement, is actually the norm (Siewierska 1999, Haig 2018b). As mentioned, clitic object pronouns are widespread across West Iranian languages. But to my knowledge, not a single language has developed obligatory object agreement based on these pronominal forms.

Note finally that object agreement *per se* is not ruled out in Iranian. A number of languages have obligatory agreement with objects in past transitive constructions, including Pashto, Northern Kurdish, or Zazaki, illustrated in (24):

- (24) *mi* *nā* *keynekī* *to-rē* *ārdā*
 1SG.OBL this girl (FEM.) 2SG-for bring.PST.FEM.SG
 ‘I brought **this girl** for you’ [Zazaki, Paul 1998: 129, glosses added]

The agreement morpheme on the verb in (24) is not etymologically related to the third person singular clitic pronouns of Table 2, or their cognates in the other languages we have been discussing. Furthermore, object agreement is primarily in the features of gender and number, rather than person. This is typical for object agreement in Iranian: where it is found, it is not etymologically related to the clitic pronouns of Table 2, and most consistently indexes the features of number, and gender, rather than person (Haig 2017, 2018b). Thus object agreement in Iranian is attested, but it has not arisen via the grammaticalization of clitic pronouns in the same manner as subject agreement has. Given the antiquity of clitic object pronouns, and their wide distribution throughout Iranian languages, the lack of object agreement derived from object pronouns is highly conspicuous, and indicative of deep differences between object and subject agreement, though obscured by the superficial similarity in form between clitic object and clitic subject pronouns.

4 Conclusions

Following changes in Iranian morphosyntax between the Old and the Middle Iranian period (perhaps 2000-2500 years ago), a paradigm of clitic pronouns (Table 2) came to be used to index past transitive subjects. In Middle Iranian, these subject clitic pronouns were in complementary distribution with free NP subjects; this kind of system is still attested in some West Iranian languages to this day. In others, the subject clitic pronouns have become fully obligatory agreement markers, illustrated for Central Kurdish in (13)-15). An identical paradigm of clitic pronouns has been used to index direct objects for even longer, and clitic object pronouns remain widespread across West Iranian to this day. But nowhere have they reached an agreement stage, despite their lack of prosodic independence and in some cases, morphological integration into the governing predicate.

The history of Iranian provides thus some support for the Minimalist account of the grammaticalization of subjects, but only the final stages in the cycle (3) are actually attested; the assumed initial stages are beyond the realms of historical attestation. We can assume that a full cycle - from emphatic pronoun to subject agreement - would involve a time span in the realm of several millennia; this would explain why an unbroken chain of attestation covering all stages of the cycle is unlikely to be forthcoming for any language. For object pronouns, however, despite the presence of a seemingly optimal configuration for the start of the cycle, there is little evidence of further developments in the predicted direction.

Why should the outcomes of the two processes have turned out differently, despite the phonological identity of the input material? It has been shown elsewhere (Haig 2018b), that there is a strong typological tendency for object agreement to be conditioned, e.g. through definiteness, topicality, or animacy of the object, rather than be fully obligatory. Subject agreement, on the other hand, tends to be across-the-board obligatory. And where obligatory object agreement is attested, it is most frequently in number and gender, rather than

person. Baker (2011) provides a partial explanation for the latter tendency from a synchronic perspective, but in this chapter we are centrally concerned with the diachronic mechanisms by which pronouns (may) become agreement markers (not all agreement markers originate from pronouns). The Minimalist account of van Gelderen (2011b) assumes general principles such as the Head Preference Principle, and Feature Economy as the driving forces behind the grammaticalization of pronouns. But neither would predict any differences between subject and object grammaticalization. An alternative usage-based explanation is set out in Haig (2018b), which points to the differing informativity of subject and object indexing with regard to the feature of person: the person value of a subject index is not readily predictable, while that of an object index (cross-linguistically, it appears that upwards of 90% of objects in discourse are actually third person; see Haig 2018b: 810-812 for details). Whether this can be confirmed remains an open question, but any account of the grammaticalization of pronouns towards agreement needs to account for the fundamental differences between subject and object pronouns in this regard.

Another important point to emerge from the Iranian data is that cliticization by itself is not necessarily the start of the slippery slope towards grammaticalization into inflectional morphology. Clitics can remain just that for millennia; there is nothing inevitable in the assumed clines for the grammaticalization of pronouns (see Schiering 2005 on the independence of phonological attrition and functional grammaticalization). This appears to be particularly true of object pronouns, which are frequently prosodically weak and attach to a verbal head, even in English. But they may evidently plateau at that stage for a very long time. As Siewierska (1999) put it, object pronouns just don't "make it" to the assumed end of the grammaticalization cline. I see no compelling grounds for assuming that sporadic cases of object clitic doubling in Persian are the first stage wholesale object agreement in this language. The notable absence of such a development anywhere else in West Iranian (i.e. from clitic pronouns cognate with Table 2 above to object

agreement marker) make this a very unlikely scenario - unless one is committed to a cyclic view of the development of agreement for both subject and object pronouns.

References

- Adibifar, Shirin. 2016. "Multi-CAST Persian". *Multi-CAST: Multilingual corpus of annotated spoken texts* ed. by Geoffrey Haig & Stefan Schnell. <https://multicast.aspra.uni-bamberg.de/#persian> (accessed 25 June 2019).
- Baker, Mark. 2011. "When agreement is for number and gender but not person". *Natural Language and Linguistic Theory* 29. 875–915.
- Chomsky, Noam. 1995. *The Minimalist Program*. Cambridge: Massachusetts Institute of Technology Press.
- Corbett, Greville G. 2006. *Agreement*. Cambridge: Cambridge University Press.
- Dabir-Moghaddam, Mohammad. 2008. "On agent clitics in Balochi in comparison with other Iranian languages". *The Baloch and others* ed. by Carina Jahani, Agnes Korn & Paul Titus, 83-100. Wiesbaden: Reichert.
- Fonseca-Greber, Bonnibeth. 2000. *The Change from Pronoun to Clitic and the Rise of Null Subjects in Spoken Swiss French*. Dissertation, University of Arizona.
- Fuß, Eric. 2005. *The rise of agreement*. Amsterdam: Benjamins.
- Gelderen, Elly van. 2011a. "The grammaticalization of agreement". *The Oxford Handbook of Grammaticalization* ed. by Bernd Heine & Heiko Narrog. Online Publication Date: Sep 2012 DOI: 10.1093/oxfordhb/9780199586783.013.0039
- Gelderen, Elly van. 2011b. *The linguistic cycle. Language change and the language faculty*. Oxford: Oxford University Press.
- Givón, Talmy. 1976. "Topic, pronoun and grammatical agreement". *Subject and topic* ed. by C. Li, 151-188. New York: Academic.
- Hadjidas, Harris & Maria Vollmer. 2015. "Multi-CAST Cypriot Greek". *Multi-CAST: Multilingual Corpus of Annotated Spoken Texts* ed. by Geoffrey Haig & Stefan Schnell. <https://multicast.aspra.uni-bamberg.de/#cypgreek> (accessed 25 June 2019).
- Haig, Geoffrey. 2008. *Alignment change in Iranian languages: A Construction Grammar approach*. Berlin: Mouton de Gruyter.

- Haig, Geoffrey. 2017. “Deconstructing Iranian ergativity”. *The Oxford handbook of ergativity* ed. by Jessica Coon, Diane Massam & Lisa Travis, 465–500. Oxford: Oxford University Press.
- Haig, Geoffrey. 2018a. “Grammaticalization and inflectionalization in Iranian”. *Grammaticalization from a typological perspective* ed. by Heiko Narrog & Bernd Heine, 57-78. Oxford: Oxford University Press.
- Haig, Geoffrey. 2018b. “The grammaticalization of object pronouns: Why differential object indexing is an attractor state”. *Linguistics* 56(4). 781-818.
- Haig, Geoffrey and Širin Adibifar. In press. “Referential Null Subjects (RNS) in colloquial spoken Persian: Does speaker familiarity have an impact?” *Essays on the typology of Iranian languages* ed. by Alireza Korangy & Behrooz Mahmoodi-Bakhtiari. Berlin: Mouton de Gruyter.
- Haig, Geoffrey & Hanna Thiele. 2015. “Multi-CAST Northern Kurdish”. *Multi-CAST Multilingual Corpus of Annotated Spoken Texts* ed. by Geoffrey Haig & Stefan Schnell. <https://multicast.aspra.uni-bamberg.de/#nkurd> (accessed 25 June 2019).
- Jügel, Thomas. 2015. *Die Entwicklung der Ergativkonstruktion im Alt- und Mitteliranischen: Eine korpusbasierte Untersuchung zu Kasus, Kongruenz und Satzbau*. Wiesbaden: Harrassowitz.
- Kahnemuyipour, Arsalan. 2003. “Syntactic categories and Persian stress”. *Natural language and linguistic theory* 21. 333-379.
- Kemenade, Ans van. 2000. “Jespersen’s Cycle revisited: Formal properties of grammaticalization”. *Diachronic syntax: Models and mechanisms* ed. by Susan Pintzuk, George Tsoulas & Anthony Warner, 51–74. Oxford: Oxford University Press.
- Korn, Agnes. 2009. “Western Iranian pronominal clitics”. *Orientalia Suecana* 58. 159–171.
- Lazard, Gilbert. 2006 [1957]. *Grammaire du persan contemporain*. Nouvelle édition avec la collaboration de Yann Richard, Rokhsareh

- Hechmati et Pollet Samvelian. Téhéran: Institut Français de Recherche en Iran. [originally published under the same title 1957]
- Lecoq, Pierre. 1979. *Le dialecte de Sivand*. Wiesbaden: Reichert.
- MacKenzie, David. 1961. *Kurdish dialect studies*, vol. 1. Oxford: Oxford University Press.
- MacKenzie, David. 1962. *Kurdish dialect studies*, vol. 2. Oxford: Oxford University Press.
- MacKenzie, David. 1966. *The dialect of Awroman (Hawrāmān-ī Luhōn). Grammatical sketch, texts, and vocabulary*. København: Det Kongelige Danske Videnskabernes Selskab.
- Öpengin, Ergin. 2013. *Clitic/affix interactions: A corpus-based study of person marking in the Mukri variety of Central Kurdish*. Ph.D. thesis, Paris III/Bamberg.
- Öpengin, Ergin. 2016. *The Mukri variety of Central Kurdish: Grammar, texts and lexicon*. Wiesbaden: Reichert.
- Öpengin, Ergin. Forthcoming. "Accounting for clitic and affix combinations in Central Kurdish". *Current Issues in Kurdish Linguistics* ed. by Gündoğdu, Songül, Ergin Öpengin, Geoffrey Haig & Erik Anonby. Bamberg: University of Bamberg Press.
- Öpengin, Ergin & Masoud Mohammadirad. To appear. "Formal and functional variation in pronominal clitics across Kurdish: documentation and explanations". *Structural and Typological Variation in the Dialects of Kurdish* ed. by Yaron Matras, Ergin Öpengin & Geoffrey Haig. London: Palgrave Macmillan.
- Paul, Ludwig. 1998. *Zazaki. Grammatik und Versuch einer Dialektologie*. Wiesbaden: Reichert.
- Rasekh, Mohammad. 2014. "Persian clitics: doubling and agreement". *Journal of Modern Languages* 24(1). 16–33.
http://umijms.um.edu.my/filebank/published_article/7262/Rasekh%20-%20JML24%20-%202016-33.pdf (accessed 10 October 2017).

- Samvelian, Pollet. 2007. “What Sorani Kurdish absolute prepositions tell us about cliticization”. *Proceedings of the Texas Linguistic Society IX: The morphosyntax of understudied languages* ed. by Frederick Hoyt, Nikki Seifert, Alexandra Teodorescu & Jessica White, 265–285. Stanford, Calif.: CSLI online publications.
http://web.stanford.edu/group/cslipublications/cslipublications/TLS/TLS9-2005/TLS9_Samvelian_Pollet.pdf (accessed 10 October 2017).
- Schiering, René. 2005. *Cliticization and the evolution of morphology: a cross-linguistic study on phonology in grammaticalization*. PhD Thesis, University of Konstanz.
- Schnell, Stefan. 2018. “Whence subject-verb agreement? Investigating the role of topicality, accessibility, and frequency in Vera’a texts”. *Linguistics* 56.4. 735-780.
- Siewierska, Anna. 1999. “From anaphoric pronoun to grammatical agreement marker: Why objects don't make it”. *Folia Linguistica* 33(1–2). 225–251.
- Siewierska, Anna. 2004. *Person*. Cambridge: Cambridge University Press.
- Stilo, Don. 2018. “The Caspian region and South Azerbaijan: Caspian and Tatic”. *The languages and linguistics of Western Asia. An areal perspective* ed. by Geoffrey Haig & Geoffrey Khan, 659-824. Berlin: Mouton de Gruyter.
- Torres Cacoullous, Rena & Catherine Travis. In press. “Variationist typology: Shared probabilistic constraints across (non-)null subject languages”. *Linguistics*.