Geoffrey Haig*

The grammaticalization of object pronouns: Why differential object indexing is an attractor state

https://doi.org/10.1515/ling-2018-0011

Abstract: While the grammaticalization of person agreement is a widely-cited and apparently uncontroversial topos of grammaticalization theory, the striking differences in the outcome of subject pronoun, and object pronoun grammaticalization, remain unexplained, and the relevant literature continues to assume a unified grammaticalization pathway. This paper argues that the grammaticalization of object pronouns is fundamentally different to that of subject pronouns. More specifically, although object pronouns may be rapid early grammaticalizers, often losing prosodic independence and cliticizing to a verbal head, they do not advance further to reach the stage of obligatory agreement markers typical of subject agreement. Typically, object markers remain at the stage of Differential Object Indexing, where their realization is conditioned by a bundle of semantic and pragmatic factors exhibiting close parallels to those operative in Differential Object Marking. Evidence from language typology, and from the diachrony of person markers across two millennia of Iranian languages, is adduced to back up these claims. Thus the widely-assumed grammaticalization cline for the grammaticalization of agreement needs to be reconsidered; for object agreement, there is evidently an attractor state, that of Differential Object Indexing, beyond which object agreement seldom proceeds. Finally, explanations grounded in discourse data are proposed, which also account for why obligatory object agreement in the category of person is so rare, while gender and number agreement for objects is far less constrained.

Keywords: grammaticalization, agreement, indexing, subject, object

***Corresponding author: Geoffrey Haig,** Allgemeine Sprachwissenschaft, Obere Karolinenstraße 8, Otto-Friedrich-Universität Bamberg, Bamberg 96049, Germany,
E-mail: geoffrey.haig@uni-bamberg.de

1 Introduction: Siewierska's puzzle

In a number of languages, independent personal pronouns exhibit striking phonological similarities to the corresponding verbal person agreement affixes. By way of illustration, consider the data in Table 1 from Iatmul (Papuan, Ndu; Sepik River, Jendraschek 2012: 141, 170), and Uzbek (Turkic; Uzbekistan, Sjoberg 1963), for free subject pronouns and the corresponding verbal affixes.[1]

Table 1: Free pronouns and bound person markers (subjects) in Iatmul and Uzbek.

	Iatmul		**Uzbek**	
	Free Pronoun	**Verbal Agreement Suffix**	**Free Pronoun**	**Verbal Agreement Suffix**
1sg	*wun*	*-wun*	*men/man*	*-man*
2sg	*mi'n*	*-mi'n*	*sen/san*	*-san*
3sg	*di'*	*-di'*	*u(l)*	*-ø*
1pl	*ni'n*	*-ni'n*	*biz*	*-miz*
2pl	*guk*	*-guk*	*siz*	*-siz*
3pl	*di'*	*-di'*	*ular*	*-lar*

Although phonological identity between the free and bound paradigms, as in Iatmul, is rare, the phonological similarities of the type illustrated by Uzbek are crosslinguistically too widely attested to be the result of mere accidental homophony. Explanation for the similarities between free pronouns and agreement affixes are generally framed in terms of a grammaticalization process: The originally free pronouns gradually coalesce 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. In the meantime, the corresponding free forms either continue to function as free pronouns, or are replaced by innovated pronouns. The process is often modeled in the form of a cline of form types, as in (1) from Fuß (2005: 4):

1 Table 1 provides only the masculine forms for Iatmul second and third person singular. The Uzbek verbal affixes are what Sjoberg (1963) refers to as "Set 1" person markers.

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

The grammaticalization account of the emergence of agreement has remained dominant in both functionalist and generative approaches to 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, while Van Gelderen (2011) sees both as instantiations of a universal "linguistic cycle". Schnell (this issue) and Bogolomova (this issue) provide further detailed case studies of different aspects of the process, in Oceanic and Nakh-Daghestanian languages respectively.

Most previous research has focused on the grammaticalization of subject pronouns (I will use "subject" here to refer to the fusion of S/A roles, except where language-specific details of morphosyntax make it necessary to draw a distinction). However, object agreement is also crosslinguistically attested (see below). In the relevant literature, it is generally assumed that the grammaticalization of object pronouns essentially follows the same path as that of subject pronouns. Bresnan and Mchombo (1987: 777) suggest 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). Siewierska (1999), however, points to an empirical fact that is bluntly at odds with the assumption of a unified grammaticalization pathway: Crosslinguistically, fully obligatory person agreement (see below for definitions) is vastly more frequent for subjects than for objects. After considering a number of possible explanations for the discrepancy between subject and objects in this respect, Siewierska finds none to be truly compelling, concluding: "I offer them in the hope that my speculations will ignite some interest in why the last stages of such a widely assumed grammaticalization process as the development of agreement markers from anaphoric pronouns are so rarely attested synchronically and why the attested instances involve A's but not O's [direct objects, P henceforth, GH]." In her more recent work (Siewierska 2004), the issue is not re-opened, and as far as I am aware, to date no serious efforts have been made to solve Siewierska's puzzle; this paper is an attempt to do so.

Existing literature largely remains silent on Siewierska's puzzle. According to earlier versions of Mainstream Generative Grammar (Culicover and Jackendoff 2005), accusative case marking on objects would have represented a potential obstacle to the development of object agreement, because

agreement was supposed to correlate with nominative case marking. However, on the Minimalist approach advocated by Fuß (2005: 85), the assumption of a "direct relation between Case checking and agreement" is rejected, and an alternative is proposed in terms of feature matching between the verb and the closest DP in the derivation, the object. Thus nothing in this framework would actually inhibit the emergence of object agreement, and in fact it would seem to facilitate it. When it comes to the mechanisms involved, Fuß (2005: 132) refers to Givón's explanation (e.g. 1976; 1979) in terms of erstwhile stylistic devices involving the doubling of a topical NP object through a pronominal clitic. Such a stylistic device may "lose its stylistic force (presumably due to an over-use), it may gain a wider distribution and eventually become obligatory in all contexts." But Fuß (2005) mentions no reason for why object pronoun grammaticalization would differ significantly from subject grammaticalization, though notably, the detailed case studies actually documented in Fuß (2005) all involve subject grammaticalization. Van Gelderen (2011: 86) likewise sees no grounds to assume any differences in the processes involved: "Like subject pronouns, independent object pronouns can attach to verbs to become dependent heads. They can subsequently be reanalyzed as agreement markers and disappear."

Functionalist accounts from Givón (1976) onwards relate the grammaticalization of subject agreement to the association of the subject role with the topic function (see Schnell (2012, this issue) for discussion). While the link of topicality to subjecthood is undeniable, it is also true that objects can bear topic relations (Dalrymple and Nikolaeva 2011: Ch. 6), though less frequently than subjects, so if topicality is indeed implicated, then we should actually expect differences in the grammaticalization of subject and object pronouns. Functionally oriented linguists nevertheless generally continue to assume that object pronouns, if they do grammaticalize, will traverse the same pathway that defines the grammaticalization of subject agreement. In Siewierska's influential work on person agreement (2004: 263), the the attested paucity of object agreement (see below) is accounted for in grammaticalization terms: "[...] subject agreement tends to be further on the grammaticalization path than object agreement." This is clearly not an explanation, but merely a re-statement of the facts. If one assumes that the grammaticalization of person markers is "a continuous process ongoing in all languages in all times" (Siewierska 2004: 251), then the question remains as to why subjects should outstrip objects in grammaticalization: Is it simply a matter of differing rates of grammaticalization, or are there fundamental differences in the processes concerned? And in either case, why should there be differences at all?

In this paper I will suggest that the subject/object asymmetry identified by Siewierska (1999) reflects fundamental differences in the grammaticalization pathways of subjects and objects. More precisely, as Siewierska's (1999) prescient comments already indicate, the differences are located in the latter stages of the assumed grammaticalization cline given in (1): Object pronouns readily undergo early grammaticalization to become prosodically dependent elements, clitics or affixes. But unlike subjects, they fail to achieve full person agreement status, and instead plateau at the stage of *Differential Object Indexing* (DOI), a term adopted from Iemmolo and Klumpp (2014). I will illustrate these claims with (still provisional) evidence from typology (Section 3), and from a diachronic case study covering 2500 years of the development of person agreement in Iranian (Section 4). In Section 5, I will formulate an explanation, based on an analysis of information structure, and the distribution of person forms in corpora of spoken languages. These explanations tie into another, related finding: The dispreference for object agreement applies specifically to the category of person, but does not hold for gender and number. Thus a set of hitherto unrelated observations are seen to be amenable to the same explanations, suggesting that we are dealing with a more general principle that shapes the way agreement systems develop historically, and consequently, how they are distributed crosslinguistically. Section 6 summarizes the main points of the paper.

2 Theoretical preliminaries

The term *agreement* has been defined in many ways; see Cysouw (2011) for a historical review. A maximally broad extension of the term is found in Siewierska (2004), who introduces a typology of different agreement types relevant for the feature of person. In this paper, I follow Haspelmath (2013) in using the term *indexing* to cover most of what Siewierska (2004) calls "agreement", while restricting my use of "agreement" to one narrowly defined type of "indexing" (close to, but not identical with, Corbett's (2006) notion of "canonical agreement"). In the current usage, indexing refers to the relationship between a NP, the controller, and an index located outside the controller, but within a specifiable syntactic domain. The index replicates a particular value for some feature of the controller, most typically person, number, and gender. An index may be prosodically independent, for example a free pronoun, or it may be prosodically dependent on a host item, in which case it is an affix, or a clitic. I refer to the former as free indices, the latter as bound indices. Thus in the

German example (2), *Arzt* is the controller, the suffix *-t* is the bound index, replicating the features of person and number of the controller:

(2) *Der* *Arzt* *sprich-t* *Italienisch*
DEF.M.SG doctor speak.PRS-3SG Italian
'The doctor speaks Italian.'

I will reserve the term *agreement* for those indices whose presence is obligatory, that is, which are required by a specific morphosyntactic configuration. In the German example above, the person index is an example of agreement, because in German all finite verbs in the present indicative require a specification for person and number. Thus speakers exercise no choice here, and an appropriate person agreement marker is always required, regardless of semantic or pragmatic considerations. This definition of agreement coincides with that of De Cat (2005), or Fuß (2005: 132): Agreement "is always obligatory [...] independent of the context".

Agreement is thus a special case of indexing, namely obligatory indexing. Note that this definition is oblivious to the presence or absence of the controller in the relevant syntactic domain (Haspelmath's (2013) "conominal", or Corbett's (2006) "multi-representation"). Obviously, a language like German which requires obligatory indexing is likely to license the co-occurrence of the controller as a free NP, though the degree to which controllers are overtly expressed is a matter of gradual crosslinguistic variation, outside the scope of this paper (see Bickel 2003; Holmberg 2009; Schnell (this issue), on the crosslinguistic variation in Referential Null Subjects, or, more generally, Referential Density). Agreement thus defined falls under the common conception of inflection, as morphology that is "obligatory in a given morphosyntactic context" (Norde 2009: 153). To claim that agreement is inflection does not preclude the possibility of agreement contributing to reference, but that is simply orthogonal to the definition of agreement adopted here. Finally, in what follows I will be largely concerned with the feature of person, and the syntactic domain of argument–predicate relations.

3 Subject and object indexing crosslinguistically: evidence from typology

Unfortunately, none of the available surveys differentiates between non-obligatory and obligatory indexing (my agreement), rendering direct

comparisons difficult. Siewierska (2013) surveys the presence vs. absence of "verbal person marking" with A and/or P in transitive clauses, based on a sample of 378 languages. Of the 290 languages that have some kind of indexing, 73 have A-agreement only, while 24 have P-agreement only. The rest (193) have both A and P agreement. Thus the tendency is that if only one argument of a transitive verb shows agreement, this will most commonly be the subject.[2]

Universal no. 293 of the Konstanz Universals Archive formulates the relationship between subject and object agreement as an implicational universal: "IF the verb agrees with the direct object, THEN the verb agrees with the subject."[3] However, this statement requires a rider: It holds only for agreement in the feature of person. In gender and number, there are counter-examples, e.g. from Nakh-Daghestanian languages (cf. Forker this issue, and Nichols this issue) and more generally languages with morphological ergativity; I return to this below. But even when restricted to person, as just suggested, it admits counter examples, such as Savosavo (Papuan, Solomon Islands, Wegener 2008). Nevertheless, it captures what seems to be a widely-held generalization regarding the relative markedness of object agreement in comparison to subject agreement: Kibrik (2011) refers to the privileged role of "principal" (=S/A) in agreement, while Bickel et al. (2013:33) note a strong statistical bias towards S=A in agreement in the majority of genetic groupings worldwide.

Object agreement is nevertheless apparently widely attested, though by all accounts significantly less frequently than subject agreement. However, none of the surveys mentioned above explicitly distinguish obligatory from non-obligatory indexing, and in fact what counts as agreement varies considerably from one study to the next. As illustration of the problems involved, consider Goldstein's (2014) account of object indexing in Lycian (Anatolia, Indo-European). The author analyzes the system as "object agreement", apparently rare within Indo-European, based on the following criteria: (i) The bound indices bear no phonological resemblance to existent free pronouns; (ii) they fail to register gender and number of the objects; (iii) they may co-occur with an overt free object ("clitic doubling"). Although these factors are certainly relevant (and are frequently invoked in discussions of agreement), they do not imply obligatoriness, and as later becomes apparent in Goldstein's analysis, the Lycian object indices are not obligatory: They do not occur when the object is indefinite, or when it is accompanied by a demonstrative. Thus the presence of object

2 This tendency is undoubtedly weaker in languages with morphological ergativity, for obvious reasons: the P-category is united with the S-category, hence the agreement relation may link P with a subset of subjects. However, the relevant figures are not available to me.

3 See http://typo.uni-konstanz.de/archive/intro/index.php (accessed 12 March 2015).

indexing in Lycian is, despite heavy phonological attenuation and prosodic dependency of the indices, determined by pragmatic features of the object NP, rather than being obligatory. Lycian thus turns out to be an unremarkable case of what I call Differential Object Indexing (see below), but this fact only emerges from a more detailed investigation of the data. Just how many languages analyzed as having "object agreement" in the literature are of this type is difficult to ascertain. Among the cases of "object agreement" listed in Siewierska (2013), a quick glance suggests that Anejom (cf. Lynch 2000), Roviana (cf. Siewierska (1999: 236), Yimas (cf. Foley 1991: 232–233), Canela-Kraho (cf. Popjes and Popjes 1986: 175–177), Macushi (Abbott 1991: 24) and Panjyima (Dench 1991: 159) are Differential Object Indexing, rather than obligatory agreement, and I assume that a systematic investigation would yield many more examples.

In sum, previous research converges on the view that subject agreement in person is crosslinguistically significantly more frequent than object agreement. But direct comparison is hampered by inconsistent definitions of "agreement". If the criterion of obligatoriness is considered, I predict that the global differences in the frequency of subject versus object indexing will turn out to be much greater than the available surveys suggest. While obligatory person indexing (agreement) for subjects is globally attested, across all major genetic groupings, fully obligatory object indexing of person is a very rare phenomenon, as Siewierska (1999) had already suspected. Although the typological verdict on this issue has yet to be passed, I suggest (3) as the provisional formulation of a testable hypothesis for future large-scale investigation:

(3) Fully obligatory agreement in the feature of person is crosslinguistically overwhelmingly more frequent, and has a less restricted areal and genetic distribution, for subjects than for objects. This tendency is weakened when:
  (a) the same object indices also index the S-role (i.e. indexing morphology groups S and P together in an ergative-type agreement relation),
  (b) the features of gender and/or number are also involved.

3.1 Differential object indexing (DOI) in the category of person

The term *Differential Object Indexing* (DOI) is adopted from Iemmolo and Klumpp (2014). It refers to a system of object indexing which applies to only a sub-set of direct objects (P), rather than across-the-board obligatory indexing of

all objects. Related phenomena have been discussed under various terms ("suspended agreement", in Bickel et al. (2007); "conditioned agreement" in Haig 2013). The term *DOI* is chosen here because it is based on the neutral term *indexing* rather than *agreement*, and because it evokes a parallelism to the better-known phenomenon of Differential Object Marking (DOM). As will become apparent, there are good reasons to postulate a link between DOI and DOM.

Differential indexing for subjects is also attested, for example in Yagua (Peba-Yaguan, Peru and Bolivia, Payne 1993). In Yagua, S_a and A arguments are indexed on the verb only if the coreferential NPs are either unexpressed, or occur after the verb. If the S_a or A argument occurs overtly, and preverbally, no index may appear on the verb (Payne 1993: 17–19). Similar systems are attested in several related Amazonian languages (Bickel and Nichols 2007), but in comparison to the attested cases of differential indexing for objects, it appears to be globally more restricted, though again this awaits further research.[4]

Differential Object Indexing in the category of person is defined as follows:

(4) Definition of Differential Object Indexing (DOI) in the category of person:
DOI refers to a system of indexing of direct objects, whereby the presence of a person index on the governing predicate is dependent on pragmatic and semantic factors outside the predicate itself. Typical conditioning factors are:
   a. the presence/absence of an overt coreferential direct object in the local clause domain,
   b. person, animacy, and information status conditions pertaining to the direct object,
   c. the information structure configuration of the entire clause,
   d. the presence of an additional object that is higher-ranked in terms of animacy.

4 Of course many languages have systematic paradigmatic gaps in subject–person agreement. Thus Russian verbs entirely lack person agreement in past tenses, while in Arabic only verbs can inflect for person, so clauses with non-verbal predicates lack person agreement affixes. English lacks any person agreement in the past tense, certain German modal verbs are also defective for person, as are converbs in e.g. Turkic, and so on. But these are systematic gaps linked to specific predicate classes, or TAM paradigms, or individual verbal lexemes, rather than to semantic/pragmatic features of individual clauses, and I would therefore not treat them as cases of Differential Subject Indexing. One phenomenon that might be worth considering as Differential Subject Indexing, however, concerns indefinite third person plural subjects in Persian, which optionally fail to trigger expected plural agreement on the verb (the verb then inflects for third singular). Note, however, that this affects number, not person.

In other words, the realization of the P index is not solely determined by a specific morphosyntactic configuration, e.g. 'finite verb form with a specific TAM-marker', but is subject to discourse pragmatic and semantic conditions pertinent to the verbal arguments. Some conditioning factors in DOI are briefly illustrated below. Turning to (4a), indexing may be sensitive to the presence vs. absence of a free object NP within a definable syntactic domain, normally the canonical object position. I will refer to this as *alternating indexing*. This is illustrated for Chichewa, in (5)–(6), where *zi-* is the bound subject index, and *–wá-* the bound object index. In (5), subject and object occur as full NPs in the clause, both in their canonical positions:

(5) *Njûchi* ***zi****-ná-lu-ma* *alenje*
bees 3PL:A-PST-bite-INDIC hunters
'The bees bit the hunters.'
(Chichewa, Bresnan and Mchombo 1987:744)

In (5), we observe the bound subject index on the verb, but a bound object index is ungrammatical, because the coreferent object NP is present in its canonical object position (see Bresnan and Mchombo 1987 for a detailed exposition of what constitutes the canonical object position). But if the object NP is either completely absent, or in a non-canonical position, it may be indexed on the verb through the bound object index *-wá-*, as in (6). The subject index, on the other hand, is obligatory, regardless of the presence, absence, or position of a subject NP:

(6) *Alenje* ***zi****-ná-****wá****-lu-ma* *njûchi*
hunters **3PL:A**-PST-**3PL:P**-bite-INDIC bees
'The hunters, the bees bit them.'
(Chichewa, Bresnan and Mchombo 1987: 745, ex. (4c), translation added)

Besides Bantu, alternating object indexing is clearly attested in Arabic (Semitic) and Iranian (see Section 4), though it is often the case that the index sporadically occurs with objects that are outside of their canonical positions (various kinds, and degrees, of "dislocation", see e.g. Woolford 2001; Krapova and Cinque 2008).

Probably more common than simple alternating indexing are systems that include animacy, or topicality-based factors pertinent to the P (4b.). In Northern Ostyak (Uralic, Khanty), an "object that is topical" triggers object agreement, while non-topical objects do not (Dalrymple and Nikolaeva 2011: 143). Note, however, that object indexing in Ostyak and related Uralic languages is in the feature of number, not person, cf. (7):

(7) *ma tam kalaŋ-ət we:l-sə-**l**-am*
1SG dem **reindeer-PL** kill-PST-**PL:P**-1SG:A
'I killed **these reindeer**.'
Northern Ostyak (Dalrymple and Nikolaeva 2011: 142)

The "clitic-doubling" that accompanies (some) Spanish direct objects is evidently related to animacy, though there is considerable regional variation regarding the relevant factors (and accompanying controversy in the literature). Belloro (2007: 15) notes as a minimal cross-dialectal consensus that "doubling is considered obligatory across all dialects if the accusative phrase is realized by a strong pronoun", illustrated in (8), adapted from Belloro (2007: 15):

(8) *Ø/ *lo=vi a él.*
3SG.M=see.PST.1SG ADP 3SG.M
'I saw him.'

Animacy-related Differential Object Indexing is widespread in Indo-European, e.g. Greek, Rumanian, south Slavic languages, Albanian, and Iranian. Outside Indo-European, animacy-conditioned DOI is found in Tukang Besi, (Austronesian, Indonesia, Donohue 1999), and in Abkhaz, where an object index is present on the verb except when the object is non-human, singular, and overtly present in the canonical object position, Hewitt (1989: 108). Similarly, in Puma (Southern Kiranti, Bickel et al. 2007), object indexing is sensitive to the distinction between human and non-human objects. In several Bantu languages, animacy effects add additional complexities to the straightforward alternating system outlined above for Chichewa (Wald 1979; Woolford 2001). Further cases of DOI are discussed in Dalrymple and Nikolaeva (2011), Schnell and Haig (2014), and Kibrik (2011: 189).

The Neo-Aramaic dialect of Barwar (Semitic, North Iraq, Khan 2008), illustrates (4c): indexing of direct objects conditioned by the information status of the entire clause. Khan (2008: 782) concludes that object indexing in this language is conditioned by "some kind of prominence of the verbal predicate", i.e. is conditioned by the clause's "status in discourse rather than the properties of its components".

The last conditioning factor considered here is (4d), the presence of a higher-ranked argument, typically a Recipient, Benefactive, or External Possessor, which is indexed on the verb at the cost of the otherwise expected object index. I refer to this as ***slot co-optation***. This is illustrated for Warlpiri (Hale 1982: 251–252, glosses slightly modified). A 3PL direct object is normally obligatorily indexed on the auxiliary:

(9) *ngajulu-rlu kapi-rna-**jana** **karli-patu** jarnti-rni*
1s-ERG FUT-1s:A-3PL:P boomerang-PL:P trim-NPST
'I will trim **the boomerangs.**'

However, the presence of a dative-marked second object in the clause leads to the co-optation of the agreement slot by an index for the higher-ranked argument:

(10) *ngajulu-rlu kapi-rna-**rla** **kurdu-ku** karli-patu punta-rni*
1s-ERG FUT-1s:A-3s:R child-DAT boomerang-PL:P take-NPST
'I am going to take the boomerangs away **from the child**.'

A very similar phenomenon is observed in Central Kurdish (Öpengin 2012; Haig 2017).

A phenomenon related to DOI is overt detransitivization triggered by objects that are non-referential, or generic. In Yup'ik Eskimo (Mithun 2004: 11–12), this extends to clauses where an object is indefinite. Under these conditions, the object is "expressed as an oblique, in Yup'ik the ablative. The clause is then grammatically intransitive". Compare (11) with a definite object and (12) with an indefinite object:

(11) *Arnam neqa nere-a-a*
woman-ERG fish(ABS) eat-INDIC.TRANS-3SG/3SG
'The woman ate **the fish**.'

(12) *Arnar neqa-meq nere-u-q*
woman(ABS) fish-ABL eat-INDIC.INTRANS-3SG
'The woman ate **a fish**.'

Although (11) – (12) evidently involve similar factors as those found in DOI, I would prefer to keep the two phenomena distinct. The Yup'ik case involves a complete detransitivization of the clause, evident in the switch from ergative to absolutive case on the subject, and the ablative case on the object. Consequently, the verb switches transitivity classes and is thus simply incapable of inflecting for object agreement. This should be kept distinct from DOI as outlined above, which involves the omission of (potentially available) object indexing, but no categorical shift in transitivity. As Bickel and Nichols (2007) show, there are obvious commonalities between various detransitivization strategies (including, for example, object incorporation) and DOI, but I will not include them under DOI at this point.

In sum, Differential Object Indexing is crosslinguistically richly attested, e.g. in Bantu, Romance, Iranian, Arabic, Abkhaz, or Uralic. Notably, in all these languages the Differential Indexing for objects co-exists with obligatory indexing for subjects. The reverse situation, obligatory object agreement coupled with Differential Subject Indexing, is unknown to me, at least on the basis of the provisional (and Eurasian-biased) group of languages that I have investigated to date.

There are very obvious parallels between the factors behind DOI, and those involved in the more extensively researched phenomenon of DOM (Iemmolo and Klumpp 2014; Seržants and Witzlack-Makarevich 2018). While the latter has often been considered a rather "exotic" phenomenon, Sinnemäki (2014) demonstrates that crosslinguistically, DOM is actually more common than unified object marking. This finding is precisely in line with the expectations of this paper regarding object indexing: DOI is not in any sense an abnormal, or defective, or a diachronically instable type of indexing. It is the crosslinguistic norm for object indexing in the category of person. Further evidence in favor of this conclusion is adduced in the following sections.

4 The emergence of Differential Object Indexing: person indexing in Iranian

In the preceding section, I formulated a preference in the way person indexing is organized crosslinguistically: subjects tend towards obligatory indexing, objects towards differential indexing. There is a growing recognition (or re-discovery) that typological variation reflects the result of diachrony, and differences are rooted in the nature of source constructions, coupled with areal and genetic biases that shape developments of individual languages (Cristofaro 2013; Bickel 2015). With regard to person indexing, we know that bound person indexing (agreement in the loose sense) is often, though not exclusively, the outcome of a historical process of grammaticalization of free pronouns. We can therefore expect the attested differences between subject and object agreement to be reflected in the historical record. However, very few language families offer the time depth of attestation necessary to trace such processes with any degree of reliability. In this section, I trace the respective grammaticalization of both subject and object indexing in the Iranian languages, which, with around 3000 years of attestation, offer a rare opportunity for tracing the trajectory of

long-term morphosyntactic change. Despite differences in the input material, and across the individual languages, by and large the outcomes of these changes conform precisely to what was predicted in the previous section: for objects, we find Differential Object Indexing, for subjects, we find obligatory agreement. Furthermore, where historical contingencies actually yielded a situation where object indexing was initially obligatory, it has demonstrably retreated backward on the grammaticalization cline toward a situation where it is differential, providing particularly striking support for the claim that differential indexing is an attractor state for object indexing.

The first set of historical developments under consideration begins with the clitic forms of non-subject pronouns, for which cognates are identifiable in Old Iranian and Old Indic (Korn 2009: Table 7). In the Old Iranian period (around the middle of the first millennium BC), these were mobile clitics, positionally conforming to Wackernagel's Law. By the start of the Middle Iranian period some two thousand years ago, the forms for Old Iranian Accusative and Genitive/Dative clitic pronouns had syncretized to a single paradigm. The forms are provided in the first column in Table 2, while the other columns show their reflexes in selected contemporary West Iranian languages.

Table 2: Pronominal clitics in Middle and contemporary West Iranian languages (simplified, based on Haig, forthcoming).

	2000 years BP	Contemporary West Iranian languages			
	Middle West Iranian	**Persian**	**Central Kurdish (Mukri)**	**Hawrami**	**Sivand**
1SG	*=m*	*=am*	*=(i)m*	*=(ı)m*	*=em*
2SG	*=t*	*=at*	*=(i)t*	*=(ı)t*	*=et*
3SG	*=š*	*=aš*	*=ī*	*=(ı)š*	*=eš*
1PL	*=mān*	*=mān*	*=mān*	*=mā*	*=emā*
2PL	*=tān*	*=tān*	*=tān*	*=tā*	*=etā*
3PL	*=šān*	*=šān*	*=yān*	*=šā*	*=ešā*

From their earliest attestations, the clitics were used to express Adnominal Possessors, Experiencers, Benefactives and External Possessors (see Haig 2008: Ch. 3). Following the syncretism of oblique cases with the accusative, the same set of clitics was also used for pronominal direct objects, and in certain environments, for subjects. I will discuss these two developments separately, beginning with the clitic pronoun objects, then the clitic pronoun subjects.

4.1 Pronominal clitics as objects

In Old and Middle Iranian, pronominal clitics could only express the object in the absence of an overt object NP (Jügel 2015: 399). Thus the object clitics were, despite their clitic status, fairly obviously pronouns, rather than any kind of agreement. Examples of object clitic pronouns from Middle Iranian (Parthian and Middle Persian) are provided below, with the object pronouns in bold type (note the Wackernagel position of the clitics):

(13) *u=**m*** *kunēd* *nām* *'Kerdir'*
and=**1SG:P** make.PRES.3SG name Kerdir
'And (he) names **me** Kerdir [...]
(Haig 2008: 114)

(14) *čīd=**mān*** *pāyēd*
always=**1PL:P** protect.PRS.3SG
'(It) always protects **us**.'
(Haig 2008: 115)

(15) *[...]* *u=**š*** *hamēw* *bōžēnd*
[...] and=**3SG:P** always save.PRES.3PL
'(The Gods) always save **him**.'
(Haig 2008: 115)

Some two thousand years later, the reflexes of the same paradigm still occurs as clitic object pronouns in numerous West Iranian languages. A fairly typical example of the clitic pronouns in object function is (16), from modern Persian:

(16) (preceding context: I said there was a sparrow on that wire)
hālā *ne-mi-bin-am=**aš***
now NEG-IPFV-see.PRS-1S=**3SG:P**
'Now I don't see **it**.'
(Modern Persian, Roberts 2009: 256)

A detailed investigation of the syntax of clitic pronouns is available for the Mukri dialect of Central Kurdish (Northwest Iranian, West Iran, Öpengin 2016). In this variety, the clitic object pronouns are regularly incorporated into the predicate, as in the following:

(17) a. *kut=ī* *“segbāb bo* *de=**m**=guž-ī?”*
say.PST=3SG.A dog.son why INDIC=**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 183–184)

This kind of “endoclisis” (Harris 2002) is typologically rare. The clitics obviously resemble affixes, because they are morphologically integrated into the predicate. Functionally, however, the object clitic of Central Kurdish remains an alternating index, only licensed in the absence of the coreferential object. This is demonstrated in (18), where the two direct objects are expressed as free pronouns, and no corresponding clitic pronoun is allowed on the verb:

(18) a. ***emin*** *de-kāte* *wezīrī* *destī.řāstī* *šā*
1SG(P) INDIC-make.PRS.3SG Vizier-of right.hand.of Shah
ʕebās-ī *w*
Abbas-OBL and
‘(God) is appointing (lit. making) **me** the right-hand vizier of Shah Abbas and
b. ***eto*** *de-kāte* *kāłek-firoš*
2sg(P) INDIC-make.PRS.3SG melon-seller
making **you** a melon-seller.’
(Öpengin 2016, KF.118–119)

Similar systems of clitic object pronouns are found in the majority of contemporary West Iranian languages, and are soundly attested in Middle Western Iranian (Parthian and Middle Persian); see e.g. Jügel (2015) for the historical data. In terms of the formal properties of the exponents, they generally exhibit clitic characteristics such as syntactic mobility (e.g. clause-second in Old and Middle Iranian, VP-second in much of Kurdish), lack of stress, and some freedom of host selection. In some languages, they even resemble affixes in that they are closely integrated into the predicate, illustrated for Central Kurdish (17). Nevertheless, nowhere do we find anything approaching obligatory object indexing. Instead, the clitic pronouns have remained pronominal expressions of the object, in complementary distribution with free NP objects, hence not permitted if the object is otherwise present in the clause.

The only exception to this are sporadic cases of clitic doubling, noted for colloquial Persian. Van Gelderen (2011: 95–96) cites examples from linguistic

literature, suggesting that clitic doubling is linked to topical objects; in her view, this is evidence that clitic object pronouns in Persian are "developing into heads and possibly agreement markers" (Van Gelderen 2011: 97). The examples cited are dependent on very specific pragmatic conditions, and fail the usual diagnostics of agreement, such as occurrence with indefinite or focal objects, as noted by Rasekh (2014). A more representative impression of the impact of this kind of construction in natural spoken Persian can be gained from the spoken texts in Adibifar (2016), based on retellings of the Pear Story video by 29 adult Persian speakers. The corpus includes over 1000 clause units, containing 628 direct objects. Of these, 46 have clitic object pronouns. In other words, less than 10% of direct objects are expressed by clitic pronouns at all. Among those 46 cases of clitic object pronouns, a sole example involves clitic doubling:

(19) *[...] yek pesar-i āmad bā dočarxe ke yeki az*
one boy-INDEF come.PST.3SG with bike COMPL one of
ān zanbilhā=rā ***gozāšt=aš*** *ruye dočarxe=aš*
those baskets=ACC put.PST.3SG=3SG.P 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)

Clitic doubling with objects is thus at best sporadically found in colloquial spoken Persian, though we lack more detailed data from larger corpora. But apart from the scarcity of object clitic doubling, there is a more fundamental flaw in the assumption that it is the first step towards agreement: there is no evidence that the sporadic cases in modern Persian are an "innovation", as Van Gelderen (2011: 96) suggests. We simply lack direct attestation of colloquial spoken Persian prior to the twentieth century, so it is perfectly possible that the kind of sporadic clitic doubling found in modern spoken Persian has been available in colloquial registers for centuries, but never found its way into writing, just as the modern clitic doubling is seldom found in the contemporary written language. Without solid evidence of increase or spread, it seems premature to assume any kind of change in progress. Finally, as noted, I am unaware of further convincing cases of object clitic doubling in other Iranian languages, although object clitics are widespread throughout the family.

In sum, with regard to the proposed grammaticalization cline from pronoun to agreement (1), object pronouns have basically plateaued at the same stage for at least 2000 years,[5] namely as clitic pronouns, perhaps accompanied by

5 If the clitic pronouns for direct objects in Old Avestan are taken into account, then the use of clitic object pronouns arguably goes back at least 3000 years in Iranian. Note, however, that the

sporadic cases of clitic doubling as illustrated in (19), though we lack diachronic data on how this may have developed. Let us consider now the history of subject clitics over the same time span.

4.2 Pronominal clitic subjects

In the Old Iranian period, the loss of inherited finite past and perfective verb forms triggered a fundamental reorganization of the morphosyntax of past-tense constructions. The result was ergative (or non-accusative) alignments in these tenses. The details have been dealt with 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 also served as subject (A) indices. Thus the typologically unusual situation arose in which one and the same paradigm of person forms indexed the direct object (P) in present tenses, illustrated in (13)–(15), and the A in past tenses (Haig 2017). 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 object indexing, because both began with phonologically identical input material.

In Old Iranian, and well into Middle Iranian, the clitic pronouns used for the A were largely restricted to occurrence in the absence of an overt A NP. Example (20) shows a clitic pronoun A, while (21) has a NP in the A role, and no clitic pronoun (both from Middle Iranian):

(20) *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 [...]'
(Haig 2008: 124)

(21) *pas **ōšbām** oy az pidar bōxt [...]*
then **ōšbām:A** 3SG from father rescue.PST.3SG
'Then **Ošbām** rescued her from (her) father [...]'
(Jügel 2015: 410, glosses added)

pronouns considered here are the syncretized forms, which are morphologically a later development than the Old Iranian accusative clitic pronouns, though their distribution includes that of the earlier accusative forms.

Further evidence for the pronominal nature of these clitics comes from the fact that they could be omitted under the condition that their reference is recoverable from the context (cf. Jügel 2015: 400). The following Middle Iranian example shows an overt clitic pronoun for the A of the first clause, and zero for the co-referential A of the subsequent clause:

(22) 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_i killed Ardawān [...] and (he_i) took his daughter as wife.'
(Jügel 2015: 411, glossed added)

Two facts thus support an 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,[6] and second, the ability to be omitted under pragmatically felicitous conditions, in a manner comparable to the omission of a free pronoun.

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 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 syntax of the Central Kurdish clitic subject indices have been amply documented in MacKenzie (1961; 1962) and Öpengin (2016), and all recent research converges on the verdict that they are exponents of an agreement relationship (Samvelian 2007; Haig 2008; Öpengin 2016). With the subject pronouns, then, we have a clear case of the expected development from alternating indexing (cf. Middle Iranian examples 18–20), corresponding to a conditioning factor (4a) in list of conditioning factors given above for object

6 Jügel (2015: 396–399) documents a number of examples from Middle Persian where the A-clitic is doubled by an overt A in the clause. Most of these are considered to involve postposed "syntactic apposition" or "afterthoughts" to the clause-initial pronominal subject. It is unclear whether such examples represent the initial stages of grammaticalization towards subject agreement, as in a Givón-type explanation, or merely coincidental accompaniments of a development that had its own dynamics; see Schnell (this issue) for a critique of Givón's explanation in terms of topic dislocation.

indexing, to a system of unconditioned, obligatory indexing in e.g. Central Kurdish (see Haig, forthcoming for details).

The history of the A-marking clitics in Iranian can therefore be considered a textbook example of the grammaticalization of subject agreement. A number of points are nevertheless noteworthy in this development. First, we cannot actually locate a previous stage involving free (as opposed to clitic) pronouns. From the start of the developments just sketched, the Old and Middle Iranian clitic pronouns were prosodically dependent clitics, and I am unaware of any serious attempt to reconstruct Proto-Indo-European free-form predecessors of these forms. Thus the attested cline of 2000 years of grammaticalization towards agreement markers only actually covers the later stages of grammaticalization, namely from a clitic pronoun, already lacking in prosodic independence, to an agreement marker, though still in clitic garb. Second, the Iranian clitic A-pronouns have not all run through the same development attested in Central Kurdish. In many languages, they have disappeared entirely, while in others, they actually remain alternating indices, rather than true agreement (e.g. Taleshi, see Paul 2011); cf. Haig (forthcoming) for a more detailed discussion of the Iranian data.

4.3 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 2 above), and both began their careers as Wackernagel clitics. In their later developments, however, they have diverged remarkably. The object clitic pronouns have 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 in Persian mentioned above. 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.

One might argue that the Iranian case merely demonstrates a different rate of grammaticalization of subject and object pronouns: given a couple more millennia, perhaps the object pronouns will eventually catch up, and Iranian will acquire true object agreement. There is a notable tendency in the literature to interpret clitic object pronouns in this manner. Radatz (2008) interprets the clitic doubling of Romance as nascent agreement markers, paving the way for a wholesale shift

towards a head-marking type. Kibrik (2011) likewise considers Romance clitic pronouns as the thin edge of the wedge in the development towards agreement markers, and in a similar line, Charitonidis (2008) sees the clitic pronouns of Greek as precursors of a rich agreement system. Bresnan and Mchombo (1987: 777) interpret the alternating object indexing of Chichewa (cf. examples (5) and (6) above) as simply lagging behind the subject indexing, and hence predict for Bantu "the existence of true grammatical object agreement, parallel to true grammatical subject agreement". The same reasoning is echoed in Van Gelderen (2011: 100), who argues that even English shows the early stages of "an object cycle", thus is likewise heading for object agreement. She points towards a historical increase in rates of prosodic incorporation of object pronouns into verbs (reflected in modern colloquial English *pick them up* > *pick'em up*), though she notes that even modern English does not permit clitic doubling with objects. On the view that loss of prosodic independence is symptomatic for a shift from pronoun to agreement marker, this might appear plausible. One of the central claims of this paper is that prosodic weakening does not necessarily correspond to a shift towards agreement, and in fact, the two processes are logically independent. Crosslinguistically, object pronouns frequently lack prosodic independence, i.e. undergo early, 'superficial' grammaticalization. But that should not be interpreted as evidence that they are condemned to proceed to the assumed endpoint. An alternative view is that for object pronouns, an attractor state exists involving a prosodically weak index, whose realization is conditioned by the kinds of factors outlined in (4) above, i.e. Differential Object Indexing. This is the natural endpoint for the grammaticalization of object pronouns, and they may remain here for millennia. This view also accounts for the crosslinguistic predominance of Differential Object Indexing in today's languages. Before turning to explanations, I will outline a further historical development, again from Iranian languages, which has hitherto proved puzzling, but receives a natural explanation in terms of the proposals outlined here.

4.4 The de-inflectionalization of object agreement in Iranian

Parallel to the grammaticalization of the clitic pronouns just discussed, a different paradigm of agreement suffixes also played a role in the person indexing system of Iranian. As mentioned above, in Old Iranian the verb system lost its finite past-tense forms, and participles became the sole carriers of past-tense semantics. Now participles, being essentially verbal adjectives, did not originally inflect for person, but for number and gender. When used predicatively, they could be combined with a finite form of the copula, and these carried regular verbal person inflection. If the copula combined with a

participle from an intransitive verb, the copula indexed person and number of the S, and when combined with a transitive verb, the copula indexed person and number of the P, thus yielding an ergative agreement system. At this early stage, then, transitive predicates consisting of participle + copula did exhibit object (P) agreement via verbal person suffixes. Crucially, however, this type of object agreement did not originate from the grammaticalization of an object pronoun, but from the re-alignment of the inherited regular verbal person inflection carried by copulas. Middle Iranian xamples of P-agreement with the copula are given below:

(23) *ME=m* *l'sd'l* *YKTLWNt'* ***HWEnd***
because=1s:A highwayman(PL) kill:PTCPL COP.**3PL:P**
'Because I killed **the highwaymen**.'
(Middle Persian, Haig 2008: 124)

(24) *xyndg* *bwd* *hym* *'w=t'n* *dryst* *(q)yrd* ***hym***
ill become.PTCPL COP.1S:S and=2PL:A healthy do.PTCPL COP.**1SG:P**
'(I) was ill and you have cured **me**.'
(MacKenzie 1979)

In a sense, then, Middle Iranian acquired person agreement with the object "for free", as the by-product of the shifts in the verb system, rather than as the culmination of a long process of grammaticalization of object pronouns. What is of interest to us now is not how this kind of object agreement entered the language, but how it subsequently developed. Basically, four main outcomes are attested. First, the paradigm of object agreement was lost, and replaced by a system of obligatory affixal subject agreement (e.g. in Persian), through analogy with agreement morphology from intransitive verbs. Second, it has been retained in some languages, notably in Northern Kurdish, though it is subject to considerable cross-dialect variation and is lost entirely in some varieties (Haig 2017). Third, it has been lost, and past transitive verbs are basically not inflected for person, either subject or object, but for plural number of the object only (Balochi, though with some additional complications, Jahani 2015). Finally, it has been retained, but no longer as obligatory object agreement, but rather as alternating object indexing. This last development is perhaps the most remarkable, and most challenging for grammaticalization theory. The phenomenon is found with past tense transitive verbs in Central Kurdish, and is sketched below.

Two paradigms of person indices are relevant for past transitives, shown in Table 3 for the Mukri dialect (Öpengin 2016).

Table 3: Person indices in the past tense of Central Kurdish (Mukri dialect).

	Verbal Suffixes	Mobile Pronominal Clitics
1SG	-im	=im
2SG	-ī	=it
3SG	-Ø	=ī
1PL	-īn	=mān
2PL	-in	=tān
3PL	-in	=yān

The mobile clitics in the right-hand column are evidently the reflexes of the Middle Iranian pronominal clitics from Table 2 above. The verbal suffixes, on the other hand, are inherited inflectional agreement suffixes; these are the suffixes which were used to index an S, and a past-tense P in Middle Iranian, as illustrated above with the copula forms in (23) and (24). A Central Kurdish example (Mukri dialect) showing this kind of agreement marker for a past-tense S is provided in (25):

(25) *(eto) roišt-ī*
2P:S leave.PST-2P:S /*-Ø
'You left.'

When indexing the S of an intransitive verb, the presence of the suffix itself is obligatory.

The presence of the free pronoun *eto* is entirely optional and would only be included for emphatic or contrastive purposes.

The same paradigm of verbal suffixes may also index a P in the past tenses, as shown in (26) for a third person plural object:

(26) *dena de=y=kušt-**in***
otherwise IPFV=3s:A=kill.PST-**3PL:P**
'Otherwise he would kill **them**.'

Crucially, however, the verbal affixes indexing a P are not obligatory agreement markers, but alternating indices. If a P argument is overtly present in the clause, the verbal affix cannot be realized:

(27) *dena **ewān**$_i$=ī de-kušt/*-**in**$_i$*
otherwise **them:P**=3s:A IPFV-kill.PST-**3PL:P**
'Otherwise he would kill **them**.'

What we find in Central Kurdish is the degrammaticalization of ancient suffixal morphology: from obligatory suffixal agreement, it has become an alternating person index, basically pronominal and anaphoric in nature, whose presence is determined by the presence/absence of the relevant NP in the clause.

This is a clear case of degrammaticalization, arguably of the type identified by Norde (2009: 152) as "deinflectionalization": "a composite change whereby an inflectional affix in a specific linguistic context gains a new function, while shifting to a less bound morpheme type". The object indices of Central Kurdish also exhibit clear signs of a looser degree of phonological integration. Positionally, these "affixes" may actually be separated from their stem by the subject-indexing "clitic", as shown in (28):

(28) *nārd=mān-in*
send.PST=1PL.A-**3PL.P**
'We sent **them**.'
(Suleimaniye dialect of Central Kurdish, MacKenzie 1961:113, glosses added)

The object index, *-in*, has been displaced by the subject clitic *=mān*. It is now the clitic that occupies the slot immediately following the stem, where it displaces the "affix". Example (29a, b) further illustrates the working of the system, this time with a second person object. In (29a), the object is overtly present in the clause as a free pronoun *to* 'you', thus no further object index is possible on the verb. This object pronoun also provides the landing site for the subject clitic, *=mān*. In (29b), the free pronoun object has been dropped for pragmatic reasons, and is therefore now indexed on the verb through the corresponding verbal agreement suffix for the second person singular, *-ī*. However, the subject clitic *=mān* now has no other landing place in the VP apart from the verb itself. Under these conditions, it displaces the object indexing suffix from the stem-adjacent position.

(29) a. *(eme)* ***to***=*mān* *nard* *bo šar-ī*
1PL:A **2S:P**=1PL:A send:PST to city-OBL
b. *(eme)* *nard=mān* ***-ī*** ...
1PL:A send.PST=1PL:A=**2P:P** ...
'We sent you to the city.'

A final piece of evidence for the loss of inflectional status of the object affixes comes from the fact that they are frequently co-opted as indices for Benefactives, Recipients, and Addressees. These arguments, when filled by a full NP, are generally expressed through adpositional phrases, thus are distinct from direct

objects. However, under pronominalization, the structure shown in (30) is possible:

(30) *ew* *beserhāt-e-ī* *bo* *gērā-**m**-ewe*
DEM real.story-DEM1-3SG:A to narrate:PST-1SG:ADDRESS-ASP
'(He) narrated this experience to me.'

The addressee (first person singular), normally the complement of the preposition *bo*, is instead indexed through a verbal affix from the paradigm of object agreement affixes. Such affix co-optation affects only the object affixes of past transitive verbs in Central Kurdish; see Öpengin (2016).

The phenomena illustrated in this section are quite remarkable from the perspective of grammaticalization theory. Degrammaticalization of inflectional affixes is not unknown, but none of the cases discussed in Norde (2009) involve person agreement, and more often a shift from inflectional to derivational is attested; there is no consensus in the literature that this involves a reversal of a grammaticalization cline. The present case, however, involves a shift from agreement affix (obligatory), to an alternating index, i.e. a sub-type of DOI.

4.5 Differential object indexing as the attractor state for object indexing

Across the entirety of the Iranian languages and their attested histories, the grammaticalization of subject and object pronouns has yielded very different outcomes, despite identical input material. The simplest generalization is that clitic pronouns indexing subjects may readily grammaticalize to become fully-obligatory agreement markers. For object pronouns, this has never happened. Instead, they plateau at various kinds of Differential Object Indexing, most commonly as alternating indexes, with perhaps sporadic and pragmatically-marked instances of clitic doubling (cf. discussion in connection with (19) on Persian). Furthermore, we have seen that object agreement can emerge through a different source, namely the re-analysis of old participial morphology), but where that has happened, it is diachronically unstable, tends to involve number and gender, and may even de-grammaticalize towards DOI.

The simplest explanation for all these phenomena is that DOI is an attractor state for object agreement generally. The term *attractor state* is taken from the theory of Complex Dynamical Systems (Van Gelder and Port 1995), which can be taken as a model for the (non-linear) complexities of language change. The term

is used here synonymously with the widely used term basin of attraction, and refers to regions within the space of variation that a complex dynamic system, over its chronological trajectory of change, tends to settle into. From such an attractor state, further development is possible, but only under the impact of an improbable confluence of contributing factors.

The term *attractor state* is apt because it also captures the typologically rare case of de-inflectionalization noted above: an erstwhile object agreement marker actually degrammaticalizes towards DOI. In other words, the diachronic shift towards the attractor state can come from either direction, i.e. through the grammaticalization of an object pronoun, or through the degrammaticalization of object agreement that had emerged from a different source.

Notice that there is abundant evidence for the earlier stages of the grammaticalization of object pronouns: loss of prosodic independence and cliticization to a verbal host. This is routinely assumed to be diagnostic of grammaticalization. However, I claim that it should not be taken as evidence for a wholesale shift towards agreement. The Lycian case (Goldstein 2014) mentioned above is very instructive in this respect: The marker concerned is phonologically heavily reduced, actually no more than a nasal feature on a verb-final segment (the reflex of an earlier object pronoun containing a nasal segment). Goldstein takes this as evidence for agreement status, presumably on the assumption that something that is phonologically eroded to this degree must be in the final stages of the grammaticalization cycle. Yet he also shows very clearly that in fact, these indices are not obligatory, but conditioned. Despite the attrited phonological form, this turns out to be a fairly standard example of DOI. The example underscores the potential independence of phonological erosion, and advancement to inflectional status. Consider also the mobile clitic pronouns of Central Kurdish: They are evidently clitics, not affixes (for example, they exhibit freedom of host selection). Functionally, however, they are inflectional markers of person agreement (Haig 2017: 482).

There is actually good reason to consider object pronouns as excellent candidates for *early grammaticalization*. Bogomolova (this issue) finds that object pronouns occur more frequently in the immediately post-verbal position than subject pronouns in Aghul, yet it is the subject pronouns that have grammaticalized to obligatory agreement markers in closely-related Tabasaran, while the object pronouns remain in complementary distribution with object NPs. Discussing object pronouns in Vera'a (Oceanic, Vanuatu), Schnell and Barth (in press) note that objects realized as lexical NP's follow the entire verbal complex (VC), while pronominal objects obligatorily occur within the VC. Here again, we find syntactic and prosodic incorporation of the object pronoun into

the predicate, a frequent phenomenon in Oceanic. Yet these object markers show no further signs of grammaticalization into object agreement. Thus mere frequent positional co-occurrence of pronoun and host does not necessarily predict the long-term outcome of grammaticalization. The difference between subject and objects makes itself apparent in the final stages, from clitic pronoun to agreement affix.

The claim that DOI is the attractor state for the grammaticalization of object pronouns accounts for all of the above, and in the absence of alternatives, it appears to be the preferable choice. The main drawback is that it involves abandoning a widely-held and cherished credo, that of a unified path in the grammaticalization of agreement, and it involves accepting that a seemingly "messy" system such as DOI may actually be diachronically stable, and cognitively undemanding (see Section 5). Furthermore, the notion of attractor state is difficult to reconcile with the view of unidirectional clines, as it explicitly endorses developments that run counter to the assumed grammaticalization cline. However, the case of degrammaticalization illustrated for Central Kurdish in (26–27) is rare, so the strength of this evidence remains to be demonstrated through further case studies. Nevertheless, I believe that there are sufficient grounds for abandoning, or at least significantly modifying, the assumption of a unified grammaticalization cline for the grammaticalization of subject and object pronouns.

5 Explanations

If the diachronic developments do indeed involve a distinct attractor state for the grammaticalization of object indexing, then we have a partial explanation for the subject/object asymmetry in agreement currently attested in the world's languages (see section 3). But even assuming that these claims turn out to be well-founded, we are still left with the question of why such an asymmetry should exist at all. I will suggest two factors, one based on crosslinguistically robust regularities of information structure, and the other on differing rates of informativeness of the relevant indices. The latter point will turn out to yield an interesting explanation for the predominance of gender and number in object agreement, and the scarcity of person.

With regard to information structure, the association of topicality with the subject role is so widely documented as to require no further comment. Conversely, objects are considered to belong more generally to that part of a proposition involving new information. The nature of the information-structural profile of objects continues to be a matter of some controversy

(Dalrymple and Nikolaeva 2011); rather than debate the issue further here, it is more instructive to consider data from corpora of natural spoken discourse. Haig and Schnell (2016b) investigate the rates of lexical expression as opposed to reduced expression (pronominal, or zero) for transitive subjects and objects, drawing on 19 corpora of naturalistic spoken language, with a total some 25,000 clause units, from 15 different languages. The contrast between A and P is highly significant: While the mean value for lexical expression of transitive subjects (A) lies around 14%, the mean for objects (P) lies at over 50%. Now "levels of lexical expression" is at best a very coarse-grained indicator of information structure, because although new information is generally expressed lexically, a significant amount of given information is also afforded lexical expression, for example for emphasis, or contrast. A more precise measure of the difference would thus consider actual discourse status, rather than simply lexical versus non-lexical. This kind of analysis is undertaken in Arnold (2003) for spoken discourse from Mapudungen (Chile, Argentina). She notes that for transitive subjects, the percentage of informationally new NP's is less than 2% (N=161), while around 48% of object NP's (N=161) encode new information (Arnold 2003: 234). Similar results have been reported for many studies, confirming the suggestion of Du Bois (1987) that P is overwhelmingly preferred for introducing new information – at least relative to A.

If we accept that these figures are representative of crosslinguistic tendencies in spoken language, then transitive subjects and transitive objects evidently differ massively in their propensity to accommodate new referents. Consequently, we can assume that rates of pronouns in A and P roles are correspondingly skewed, though this is difficult to compare directly due to crosslinguistic differences in the tolerance of zero-expression, cf. Bickel (2003). If the grammaticalization of person agreement is dependent on pronominalization of the relevant arguments, then obviously we would expect to find differences in the grammaticalization of subjects and objects, simply due to different rates in the occurrence of the respective free pronouns. This, then, would be a fairly simplistic, frequency-driven account of the differences in grammaticalization, ultimately grounded in information structure profiles typically associated with different argument types (see Schnell (this issue), and Bogomolova (this issue), for evidence pertinent to frequency-based accounts of grammaticalization). Although I believe this is undoubtedly part of the story, I suspect it is only one of several relevant factors.

The first problem with the frequency approach is that it fails to predict the attested early grammaticalization of object pronouns. As discussed above, there is abundant evidence for prosodically dependent object pronouns (clitics) in the

languages of the world and indeed, good reasons for this early grammaticalization. First, on the assumption that linear adjacency is a precondition for grammaticalization to occur, it is notable that in the majority of languages, objects occur at least as close to a verbal head as subjects do (cf. the relative infrequency of the orders OSV and VSO in the world's languages). Second, when pronominal objects differ positionally from full NP objects, the difference is generally in terms of *increased* proximity to the verb. Consider the English example (31), where a verbal particle *off* can optionally intervene between an object NP and the lexical verb:

(31) a. *take off [your shoes]*$_{\text{Obj}}$
b. *take [your shoes]*$_{\text{Obj}}$ *off*

Under pronominalization of the object, the verbal particle *off* can no longer intervene between verb and object, and the object pronoun typically undergoes phonological reduction (32b) - often interpreted as evidence of early grammaticalization (e.g. Van Gelderen 2011: 100):

(32) a. *take [them]*$_{\text{Obj}}$ *off*
b. *take'em off*
c. ??*take off [them]*$_{\text{Obj}}$ (or with attempted cliticization: **take off'em*)

Similarly, in many Oceanic languages, pronominal objects are not prosodically independent, but are incorporated into the verbal complex. Indeed, Siewierska (2004: 42) notes that some languages lack free object pronouns entirely, but none lack free subject pronouns. In other words, when objects are pronominalized, they are frequently prosodically weak and gravitate towards a lexical head. This occurs despite the fact that objects are, as the figures above show, much less frequently pronominalized than subjects. Thus object pronouns are rapid early grammaticalizers, irrespective of their actual frequency. What needs to be explained is why, despite early grammaticalization, they subsequently "don't make it" (Siewierska 1999) to fully grammaticalized agreement markers. It is less obvious why considerations of topicality, and frequency of occurrence of free pronouns, should be relevant to this later stage of grammaticalization. To understand the differences, then, we may need to look at the nature of verbal inflection, and the categories involved.

An observation repeatedly made in this paper is that the tendency to avoid object agreement is only valid for the category of person. With regard to gender (including noun class systems) and number, object indexing

systems exhibiting the obligatory characteristics of agreement are well-documented; see Forker (this issue), for obligatory gender agreement in Nakh-Daghestanian, and Dixon (2004) for Jarawara, or Mirdehghan and Jahangiri (2005) for Pashto. Number is likewise widespread in object agreement, e.g. Ostyak, Uralic (Dalrymple and Nikolaeva 2011), where only the number feature of the object is indexed (in a DOI system, rather than as obligatory agreement). In Balochi (Iranian, Iran, Pakistan and Afghanistan), the inherited system of object agreement suffixes now only reflects number (Korn 2009). Baker (2011) likewise discusses the tendency for subject agreement to be associated with person, while object agreement tends towards gender and number. While again, a detailed typological survey is lacking, the available evidence suggests an association of obligatory agreement in the category of person (with or without gender and number) with the subject role, but in the object role, person by itself is exceedingly rare.

This tendency likewise has no natural explanation within the grammaticalization framework. After all, if the low frequency levels of object pronouns were responsible for the low levels of object agreement, there would be no obvious reason why that should not operate across the board, inhibiting grammaticalization of any kind of object agreement, whether in person, number, or gender. My tentative explanation is in terms of the relative informativeness of person, gender and number indexing for the subject and object roles respectively.

To illustrate how this works, consider the person values attested in the A and P roles in natural language data. Table 4 shows data from nine language

Table 4: Percentage of third person forms in A and P roles, in spoken language corpora.

Corpus Source	Language	Text Type	% 3rd pers.	
			A	P
Switchboard	English	conversational	36	93
Multi-CAST	Cypr. Greek	folk tales	66	94
Multi-CAST	English	oral history	43	94
Multi-CAST	Nafsan	folk tales	77	95
Multi-CAST	N. Kurdish	folk tales	67	93
Multi-CAST	Persian	Pear Stories	96	100
Multi-CAST	Teop	folk tales	74	95
Multi-CAST	Tondano	oral history, video stimuli	86	96
Multi-CAST	Vera'a	folk tales	84	91
		Mean	70	95
		SD	19.0	2.46

corpora, taken from two sources. The first set stems from Bresnan et al. (2001), drawn from the SWITCHBOARD Corpus of Conversational American English (Godfrey and Holliman 1993). Bresnan et al. (2001) analyzed the one-million-word parsed section of the corpus, and investigated different person constellations in transitive clauses, collapsing the first and second persons, and contrasting it with the third person. The other eight corpora are from the Multi-CAST archive (Haig and Schnell 2016a), each containing at least 1000 clause units. For each corpus, the percentage of third person forms among all persons in A and P roles respectively has been calculated.

The crucial point to note is the respective ranges, expressed here by the Standard Deviation (SD). In the P-role, we find third person forms consistently accounting for > 90% of all objects, and a low SD of 2.46. Note that the language-specific value is largely impervious to text type (for instance, the conversational data of the Switchboard corpus does not differ from the folk-tale data from Northern Kurdish, and so on). For the A-role, on the other hand, the respective percentage of third person forms varies massively, depending on text type. In conversational data, or oral history, first and second person subjects dominate, for obvious reasons, and we have correspondingly low percentages of third person subjects. In Pear Story retellings, on the other hand, practically no first or second person subjects occur, simply because the entire narrative recounts the actions of other people.

From this, we can assume the following with respect to person agreement: In actual usage, the category of person is relatively *uninformative* in the P role. Speakers can fairly reliably predict that around 90% of objects will be third person, and this appears to be invariant across different speech situations, and languages. If we assume that speakers are sensitive to these kinds of strong-frequency effects (see Bresnan et al. 2007 for evidence that this is the case), then the inference a speaker can draw from available input is that, all other things being equal, with around 90% probability a direct object of any given transitive verb will be third person. In other words, even in the absence of any person indexing for object (or any other cues), a speaker can predict with a reasonable degree of reliability that the object is third person. The same cannot be said of the A-role, however, where we find person values vary widely according to genre and text-type. In other words, there is no reliable association of a particular person value with the subject role. Thus, in actual language usage, person indexing on subjects is vastly more informative than person marking on objects.

Assuming for a moment that the 90% figure for third person objects is something like a valid mean for discourse crosslinguistically, it does not seem unreasonable that speakers are attuned to it, and that over time, the

development of morphosyntax will be tipped in a direction that reflects this. For a language to develop obligatory object indexing would mean, in effect, that speakers are adding a person index to every transitive verb in their language which, in around 90% of the cases, would realize the same value (third person), i.e. be fairly redundant. Thus we might assume that, given restricted resources for inflection on the verb, obligatory person indexing (agreement) would be greatly favored for subjects, rather than for objects.

Gender and number, on the other hand, are informative precisely for the third person. Given a "third person singular" index, the set of potential referents is potentially very large and in the case of competing referents, such an index may contribute very little to establishing reference. But an indication of gender, or number, may narrow down the available interpretations considerably. Basically the same tendency is reflected in Greenberg's Implicational Universal number 44: If a language has gender distinctions in the first person, it always has gender distinctions in the second or third person, or in both. In other words, crosslinguistically, the third person is the unmarked locus for gender agreement. The reason is fairly obvious: The first and second person are uniquely identifiable via their person specification, so additional information on gender is redundant. The same is not true for third person referents. My suggestion is that the same factors are operative in the preference for gender agreement in the object role. As we have seen, the overwhelming majority of objects are third person. Gender is thus most meaningful in an environment where third person referents dominate, hence object agreement in gender (and/or number) is more likely than in person.

These ideas are predicated on the assumption that robust and crosslinguistically consistent tendencies in the way semantic and pragmatic categories map onto morphosyntax will be reflected, at least indirectly, in the way morphosyntax is organized. Assumptions along these lines are at the heart of what could be termed "usage-based" approaches to morphosyntax. While the wider ramifications of the "grammar vs. usage" debate are beyond the scope of this paper, I would nevertheless point out that the present paper has invoked usage-based explanations only to the extent that they are justified by verifiable statistical tendencies in corpora of natural spoken language. If the claim made in this paper, namely that there is a diachronic tendency to inhibit object agreement in the category of person, should indeed turn out to be crosslinguistically valid, then the corresponding patterns from usage should be sufficiently powerful and immune to variability to have left their mark across different genetic groupings, and over many millennia.

6 Conclusions

This paper represents a provisional answer to "Siewierska's puzzle": If the grammaticalization of personal pronouns towards agreement is a universally attested diachronic process, why should objects so rarely make it to the stage of obligatory agreement? In answering this question, I first examined evidence from typology, and from historically-attested cases of the (de-)grammaticalization of person marking, concluding that the grammaticalization of object indexing in the category of person differs crucially from that of subject indexing: Although object pronouns are rapid early grammaticalizers, readily losing prosodic independence and fusing with verbal heads, they plateau at DOI, and do not advance further towards agreement. Thus the presence of DOI in a language, or group of languages, e.g. Bantu, is not necessarily portentous of an ongoing shift towards object agreement. Rather, the Bantu data, and similar patterns from other language families, faithfully reflects the fact that the attractor state for object indexing is conditioned (or differential), rather than obligatory indexing, hence diachronically stable, and synchronically widely attested. Although there may well be diachronic shifts in the nature of the conditioning factors involved in DOI, there is rarely, if ever, a shift to full obligatory agreement.

These findings have considerable ramifications for grammaticalization theory. The evidence from object pronouns suggests two distinct sub-processes are involved: First, an initial, fairly superficial, loss of prosodic independence, leaving object pronouns adjacent to, and often cliticizing with, verbal heads. This can be witnessed in countless languages, including English. However, this process need not be accompanied by any shift in the pronominal status of the elements involved, though obviously many languages do permit various kinds of clitic doubling, leading to systems of Differential Object Indexing. Second, a development to obligatory agreement, involving a qualitatively distinct process, which I refer to as *inflectionalization*. It is not simply 'more of the same', i.e. not simply the advancement of the initial superficial grammaticalization, but is dependent on specific and possibly quite abstract requirements, about which relatively little is known, but which are evidently readily available for exponents of the subject relation, but not the object relation. The traditional view of grammaticalization in terms of a monolithic cline from pronoun to agreement obscures this fact, and leads to the unjustified expectation that inflectionalization is basically just the continuation of superficial grammaticalization.

By way of explanation for the distinct behavior of subject and object pronouns, two accounts were explored: First, the well-known asymmetry in the frequency of pronominalization (or at least reduced expression) of transitive subjects and objects, as demonstrated by crosslinguistic corpus data. Second, the relative informativeness of the category of person in the subject role, when compared to the object role. In actual usage, objects are naturally associated with the third person, rendering person a relatively uninformative category for objects in comparison with subjects. The categories of gender and number, on the other hand, are informative as they provide a means of distinguishing among different third person referents. This explanation predicts that where object agreement is found, it will be in the categories of gender and number, rather than person.

While the claims in this paper remain provisional, they unite a number of disparate observations from historical linguistics, language typology, and corpus data from natural discourse. In the course of the exposition, I have also suggested a significant modification to one of the most regularly and uncritically repeated claims in historical linguistics, namely that of a unified cline in the grammaticalization of person agreement from personal pronouns.

References

Abbott, Miriam. 1991. Macushi. In Desmond C Derbyshire & Geoffrey K Pullum (eds.), *Handbook of Amazonian languages*, vol. 3 23–160. Berlin & New York: De Gruyter.

Adibifar, Shirin. 2016. Persian. In Geoffrey Haig & Stefan Schnell (eds.), *Multi-CAST (Multilingual corpus of annotated spoken texts)*. https://lac.uni-koeln.de/multicast-persian/ (accessed 08 January 2017).

Arnold, Jennifer. 2003. Multiple constraints on reference form: Null, pronominal, and full reference in Mapudungun. In John Du Bois, Lorraine Kumpf and William Ashby (eds.), *Preferred Argument Structure. Grammar as architecture for function*, 225–245. Amsterdam & Philadelphia: John Benjamins.

Baker, Mark. 2011. When agreement is for number and gender but not person. *Natural language and Linguistic Theory* 29(4). 875–915.

Belloro, Valeria. 2007. Spanish clitic doubling: A study of the syntax-pragmatics interface. Buffalo, NY: University of New York at Buffalo dissertation.

Bickel, Balthasar. 2003. Referential density in discourse and syntactic typology. *Language* 79 (4). 708–736.

Bickel, Balthasar. 2015. Distributional typology: Statistical inquiries into the dynamics of linguistic diversity. In Bernd Heine & Heiko Narrog (eds.), *The Oxford handbook of linguistic analysis*, 2nd, 901–923. Oxford: Oxford University Press.

Bickel, Balthasar, Giorgio Iemmolo, Taras Zakharko & Alena Witzlack-Makarevich. 2013. Patterns of alignment in verb agreement. In Dik Bakker & Martin Haspelmath (eds.), *Languages across boundaries: Studies in memory of Anna Siewierska*, 15–36. Berlin & Boston: De Gruyter Mouton.

Bickel, Balthasar & Johanna Nichols. 2007. Inflectional morphology. In Timothy Shopen (ed.), *Language typology and syntactic description, rev*, 2nd edn., 169–240. Cambridge: Cambridge University Press.

Bresnan, Joan, Anna Cuenit, Tatiana Nikitina & R. Harald Baayen. 2007. Predicting the dative alternation. In Gerlof Bouma, Irene Krämer & Joost Zwarts (eds.), *Cognitive foundations of interpretation*, 69–94. Amsterdam: Royal Netherlands Academy ofScience.

Bresnan, Joan, Shipra Dingare & Christopher Manning. 2001. Soft constraints mirror hard constraints. Voice and person in English and Lummi. In Miriam Butt & Tracy Holloway King (eds.), *Proceedings of the LFG01 Conference*, 13–32. Hong Kong: University of Hong Kong. [page numbers according to http://web.stanford.edu/group/cslipublications/cslipublications/LFG/6/pdfs/lfg01.pdf.

Bresnan, Joan & Sam A Mchombo. 1987. Topic, pronoun and agreement in Chichewa. *Language* 63(4). 741–782.

Charitonidis, Chariton. 2008. Polysynthetic tendencies in Modern Greek. *Linguistik online* 34 (2). 17–40.

Corbett, Greville G. 2006. *Agreement*. Cambridge: Cambridge University Press.

Cristofaro, Sonia. 2013. The referential hierarchy: Reviewing the evidence in diachronic perspective. In Dik Bakker & Martin Haspelmath (eds.), *Languages across boundaries: Studies in memory of Anna Siewierska*, 69–93. Berlin & Boston: De Gruyter Mouton.

Culicover, Peter & Ray Jackendoff. 2005. *Simpler syntax*. Oxford: Oxford University Press.

Cysouw, Michael. 2011. Very atypical agreement indeed. *Theoretical Linguistics* 37(3–4). 153–160.

Dalrymple, Mary & Irina Nikolaeva. 2011. *Objects and information structure*. Cambridge: Cambridge University Press.

De Cat, Cécile. 2005. French subject clitics are not agreement markers. *Lingua* 115. 1195–1219.

Dench, Alan. 1991. Panjyima. In R.M.W. Dixon and Barry Blake (eds.), *The Handbook of Australian Languages*, 124–243. Oxford: Oxford University Press.

Dixon, Robert M. W. 2004. *The Jarawara language of southern Amazonia*. New York: Oxford University Press.

Donohue, Mark. 1999. *A grammar of Tukang Besi*. Berlin & New York: Mouton de Gruyter.

Du Bois, John. 1987. The discourse basis of ergativity. *Language* 63(4). 805–855.

Foley, William. 1991. *The Yimas language of New Guinea*. Stanford: Stanford University Press.

Fuß, Erik. 2005. *The rise of agreement, a formal approach to the syntax and grammaticalization of verbal inflection*. Amsterdam & Philadelphia: John Benjamins.

Givón, Talmy. 1976. Topic, pronoun and grammatical agreement. In Charles Li (ed.), *Subject and topic*, 151–188. New York: Academic Press.

Givón, Talmy. 1979. From discourse to syntax: Grammar as a processing strategy. In Talmy Givón (ed.), *Discourse and syntax*, 81–112. New York: Academic.

Godfrey, John & Edward Holliman. 1993. *Switchboard-1 Release 2 LDC97S62*. Philadelphia, PA: Linguistic Data Consortium. Web download.

Goldstein, David. 2014. Object agreement in Lycian. *Historische Sprachforschung/Historical Linguistics* 127(1). 101–124.

Haig, Geoffrey. 2008. *Alignment change in Iranian languages: A construction grammar approach*. Berlin & New York: Mouton de Gruyter.

Haig, Geoffrey. 2013. The subject/object asymmetry in bound person indexing: Diachronic and discourse considerations. Paper presented at the Workshop Agreement in Discourse, University of Bamberg, 1–2 February.

Haig, Geoffrey. 2017. Deconstructing Iranian ergativity. In Jessica Coon, Diane Massam & Lisa Travis (eds.), *The Oxford handbook of ergativity*, 465–500. Oxford: Oxford University Press.

Haig, Geoffrey. forthcoming. Grammaticalization and inflectionalization in Iranian. In Heiko Narrog & Bernd Heine (eds), *Grammaticalization from a typological perspective*, Oxford: Oxford University Press.

Haig, Geoffrey & Stefan Schnell. 2016a. *Multi-CAST (Multilingual corpus of annotated spoken texts)*. https://lac.uni-koeln.de/multicast/(accessed 24 March 2016).

Haig, Geoffrey & Stefan Schnell. 2016b. The discourse basis of ergativity revisited. *Language* 92 (3). 591–618.

Hale, Kenneth. 1982. Some essential features of Warlpiri verbal clauses. In Stephen M Schwartz (ed.), *Papers in Warlpiri grammar: In memory of Lothar Jagst* (Series A 6), 217–314. Darwin: SIL, Australian Aborigines Branch.

Harris, Alice. 2002. *Endoclitics and the origins of Udi morphosyntax.* Oxford: Oxford University Press.

Haspelmath, Martin. 2013. Argument indexing: A conceptual framework for the syntactic status of bound person forms. In Dik Bakker & Martin Haspelmath (eds.), *Languages across boundaries: Studies in memory of Anna Siewierska*, 197–226. Berlin & Boston: De Gruyter Mouton.

Hewitt, Brian G. 1989. *Abkhaz*. London: Routledge.

Holmberg, Anders. 2009. Null subject parameters. In Theresa Biberauer, Anders Holmberg & Michelle Sheehan (eds.), *Parametric variation: Null subjects in Minimalist Theory*, 88–124. Cambridge: Cambridge University Press.

Iemmolo, Giorgio & Gerson Klumpp. 2014. Introduction. (special issue: Differential Object Marking: Theoretical and empirical issues). *Linguistics* 52(2). 271–279.

Jahani, Carina. 2015. Complex predicates and the issue of transitivity: The case of Southern Balochi. In Iván Szantó (ed.), *From Aṣl to Zā'id: Essays in honour Éva M. Jeremiás*, 79–105. Piliscsaba: The Avicenna Institute of Middle Eastern Studies.

Jendraschek, Gerd. 2012. *A grammar of Iatmul*, vol. 1. 1–11. Regensburg: University of Regensburg Habilitationsschrift.

Jügel, Thomas. 2015. *Die Entwicklung der Ergativkonstruktion im Alt- und Mitteliranischen: Eine korpusbasierte Untersuchung zu Kasus, Kongruenz und Satzbau*. Wiesbaden: Harrassowitz.

Khan, Geoffrey. 2008. *The Neo-Aramaic dialect of Barwar*, vol. 3. Leiden & Boston: Brill.

Kibrik, Andrej. 2011. *Reference in discourse*. Oxford: Oxford University Press.

Korn, Agnes. 2009. Western Iranian pronominal clitics. *Orientalia Suecana* 58. 159–171.

Krapova, Iliyana & Guglielmo Cinque. 2008. Clitic reduplication constructions in Bulgarian. In Dalina Kallulli & Liliane Tasmowski (eds.), *Clitic doubling in the Balkan languages*, 257–287. Amsterdam & Philadelphia: John Benjamins.

Lynch, John. 2000. *A grammar of Anejom*. Canberra: Research School of Pacific and Asian Studies.

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. 1979. Mani's šāBUHRAGāN. *BSOAS* 42(3). 500–534.

Mirdehghan, Mahinnaz & Nader Jahangiri. 2005. Split ergative morphology in Hindu/ Urdu, Pashto, and Balochi languages. *Journal of Humanities* 12(3). 93–122.

Mithun, Marianne. 2004. Functional perspectives on syntactic change. In Brian Joseph & Richard Janda (eds.), *The handbook of historical linguistics*, 552–572. Malden, MA: Blackwell.

Norde, Muriel. 2009. *Degrammaticalization*. Oxford: Oxford University Press.

Öpengin, Ergin. 2012. Adpositions and argument indexing in the Mukri variety of Central Kurdish: Focus on ditransitive constructions. *Orientalia Suecana* 61. 187–198.

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

Paul, Daniel. 2011. *A comparative dialectal description of Iranian Taleshi*. Manchester: University of Manchester dissertation.

Payne, Thomas. 1993. *The twins stories: Participant coding in Yagua narrative*. Berkeley, CA: University of California Press.

Popjes, Jack & Jo Popjes. 1986. Canela-Krahô. In Desmond C Derbyshire & Geoffrey Pullum (eds.), *Handbook of Amazonian languages*, vol. 1 128–199. Berlin & New York: Mouton de Gruyter.

Radatz, Hans-Ingo. 2008. Non-lexical core-arguments in Basque, German and Romance: How (and why) Spanish syntax is shifting towards clausal headmarking and morphological cross-reference. In Ulrich Detges & Richard Waltereit (eds.), *The paradox of grammatical change: Perspectives from Romance*, 181–215. Amsterdam & Philadelphia: John Benjamins.

Rasekh, Mohammad. 2014. Persian clitics: Doubling and agreement. *Journal of Modern Languages* 24(1). 16–33.

Roberts, John R. 2009. *A study of Persian discourse structure*. Uppsala: Acta Universitatis Uppsaliensa.

Samvelian, Pollet. 2007. What Sorani Kurdish absolute prepositions tell us about cliticization. In Frederick Hoyt, Nikki Seifert, Alexandra Teodorescu & Jessica White (eds.), *Proceedings of the Texas Linguistic Society IX: The morphosyntax of understudied languages*, 265–285. Stanford, CA: CSLI. http://web.stanford.edu/group/cslipublications/cslipublications/TLS/TLS9-2005/TLS9_Samvelian_Pollet.pdf (accessed 10 October 2017)).

Schnell, Stefan. 2012. Explaining formal variation in subjects and objects in Vera'a: The emergence of subject-TAM markers. Paper presented at the conference New Ways of Analyzing Variation, Asia – Pacific 2, Tokyo, 1–4 August.

Schnell, Stefan & Danielle Barth. in press. Discourse motivations for pronominal and zero objects across registers in Vera'a. *Language Variation and Change* 30. 1.

Schnell, Stefan & Geoffrey Haig. 2014. Assessing the relationship between object topicalisation and the grammaticalisation of object agreement. In Lauren Gawne & Jill Vaughan (eds.), *Selected papers from the 44th Conference of the Australian Linguistic Society, 2013*, 102–125. Melbourne: University of Melbourne. http://hdl.handle.net/11343/40959 (accessed 10 October 2017)).

Seržant, Ilya A & Alena Witzlack-Makarevich. 2018. Differential argument marking: Patterns of variation. In Ilja A Seržant & Alena Witzlack-Makarevich (eds.), *Diachrony of Differential Argument Marking* (Studies in Diversity Linguistics), 1–40. Berlin: Language Science Press. doi: 10.5281/zenodo.1228243

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.

Siewierska, Anna. 2013. Verbal person marking. In Matthew S Dryer & Martin Haspelmath (eds.), *The world atlas of language structures online*, Leipzig: Max Planck Institute for Evolutionary Anthropology. http://wals.info/chapter/102 (accessed 12 January 2017)).

Sinnemäki, Kaius. 2014. A typological perspective on Differential Object Marking. *Linguistics* 52 (2). 281–313.

Sjoberg, Andrée F. 1963. *Uzbek structural grammar* (Uralic and Altaic Series 18.). Bloomington, IN: Indiana University Press.

Van Gelder, Timothy & Robert F. Port. 1995. It's about time: An overview of the dynamical approach to cognition. In Robert F Port & Timothy Van Gelder (eds.), *Mind as motion: Explorations in the dynamics of cognition*, 1–43. Cambridge, MA: MIT Press.

Van Gelderen, Elly. 2011. *The linguistic cycle: Language change and the language faculty*. Oxford: Oxford University Press.

Wald, Benji. 1979. The development of the Swahili object marker: A study of the interaction of syntax and discourse. In Talmy Givón (ed.), *Discourse and syntax*, 505–524. Amsterdam & Philadelphia: John Benjamins.

Wegener, Claudia. 2008. *A grammar of Savosavo, a Papuan language of the Solomon Islands*. Nijmegen: MPI for Psycholinguistics, Radboud Universiteit Nijmegen dissertation.

Woolford, Ellen. 2001. Conditions on object agreement in Ruwund (Bantu). In Elena Benedicto (ed.), *The UMass volume on indigenous languages* (University of Massachusetts Occasional Papers in Linguistics 20), 177–201. Amherst, MA: GLSA.