

Cognitive pressure, areal contact, or both? An evolutionary study of classifiers and number marking in Indo-Iranian

Chundra Cathcart¹, Andreas Hölzl¹, Gerhard Jäger², Paul Widmer¹, and Balthasar Bickel¹

¹Department of Comparative Language Sciences, University of Zurich

²Department of Linguistics, University of Tübingen

September 1, 2019

Abstract

It is frequently hypothesized that the presence of sortal numeral classifiers is negatively correlated with obligatory number marking on nouns. While the exact cognitive mechanisms that cause this association (known as the Greenberg-Sanches-Slobin or GSS generalization) are still debated, the hypothesis predicts an effect on language change: classifiers are predicted to develop more frequently in languages with optional plural marking than in languages with obligatory plural marking. In this paper, we test this prediction in Indo-Iranian using a Bayesian phylogenetic model. We infer posterior distributions over evolutionary transition rates between typological states. We then take these rates to reconstruct the history of classifiers and number marking throughout the family, constrained by historically attested states. We find partial but not thorough support for the GSS hypothesis. Some instances of classifier development appear to be influenced by the presence of optional plural marking, while others do not. A survey of the matter and pattern borrowing that has taken place in these languages suggests that in both cases, language contact may have played a role. Taken together, these findings tentatively suggest that the association of classifiers and optional number marking in Indo-Iranian is neither solely the effect of universal mechanisms nor of the contingency of local contact histories, but that the association was driven by the interaction of the two factors: classifiers are likely to have spread easiest where obligatory number marking has declined.

1 Introduction

Indo-Iranian languages display considerable diversity in constructions where items are enumerated. While attested ancient (and to some extent medieval) Indo-Iranian languages show a straightforward pattern where numerals or quantifiers and enumerated nouns agree fully in gender and number (thanks to the rich morphology of the older languages), contemporary languages do not consistently mark plural number on semantically plural nouns. Additionally, a number of Indo-Iranian languages make use of sortal numeral classifiers, as in the Bengali example *chɔ-ʈa boi* ‘six books’, where *ʈa* is a classifying element that co-occurs with an enumerated entity.

This behavior is typologically uncharacteristic of the larger Indo-European family and it has therefore been attributed by some researchers to contact with languages from other stocks (Emeneau 1956, 1980, Matisoff 1978, Thomason and Kaufman 1988). Others see Indo-Iranian numeral classifiers as the grammaticalized outcome of a general tendency seen in ancient and medieval Indo-European languages where generic and non-generic nouns are placed together in close apposition, which can potentially lead to systems of nominal classification (Hackstein 2010). An important confound in resolving this debate is the extent to which languages are subject to what is known as the Greenberg-Sanches-Slobin generalization (Greenberg 1972, Sanches and Slobin 1973). This generalization posits an association between the presence of numeral classifiers in a language with optionality (or even absence) of plural marking. If numeral classifiers develop to a stronger extent in languages with optional than in languages with obligatory number marking, their emergence in Indo-Iranian might reflect a general tendency of the sort argued for in the Greenberg-Sanches-Slobin hypothesis. If not, Indo-Iranian classifiers may have arisen due to contingencies of the family’s history, especially its contact history.

This paper explores the diachronic pathways by which these diverse patterns have developed, employing an explicitly phylogenetic approach and operationalizing the Greenberg-Sanches-Slobin generalization by means of evolutionary transition rates between typological states (concerning the presence of numeral classifiers and optionality of plural marking), which are estimated from the data. We find partial but ultimately incomplete evidence for the Greenberg-Sanches-Slobin generalization, indicating that both cognitive and areal pressures have played a role in shaping Indo-Iranian classifier systems, and identify likely points of language contact.

In what follows, we describe the Greenberg-Sanches-Slobin generalization in more detail (Section 2). We then provide a detailed description of the synchronic and diachronic patterns seen in Indo-Iranian constructions where items are enumerated (Section 3). After introducing our data coding and the inference model (Section 4), we present and discuss our main findings (Sections 5–7).

2 Background

2.1 Numeral Classifiers

Historically, numeral classifiers have also been referred to as *numeratives*, because they are usually contiguous to numerals in expressions of quantity (Aikhenvald 2000:98) or, more generally, found to occur in the context of quantification (Grinevald 2000:63). In this study, we define a numeral classifier as any morpheme that, independent of its morphosyntactic status, is linearly adjacent to a numeral (or an equivalent quantifier) when it occurs, functioning with the numeral as an attribute of a head noun. A numeral classifier tends to have mutual dependencies (e.g., collocational, morphosyntactic, phonological, or semantic) with the numeral and the head noun. In the following example of Mandarin Chinese, for instance, the noun *shū* ‘book’ can only be combined with the numeral classifier *běn*. Furthermore, the numeral *liǎng* replaces the general numeral *èr* ‘two’ in numeral classifier constructions. If *běn* is not preceded by a numeral it has a set of different meanings.

(1) Mandarin (Sino-Tibetan)

liǎng běn shū
two.CLF CLF book
‘two books’

Numeral classifiers usually form a constituent with the numeral rather than with the head noun (e.g., Aikhenvald 2000:105); the noun can never occur in between the numeral and the classifier (Her 2017).

Numeral classifiers are generally divided into mensural and sortal subtypes. Mensural classifiers aid in partitioning an uncountable noun (e.g., Modern German *ein Glas Bier* ‘a glass of beer’), and are usually considered to be distinct from pseudo-partitives (e.g., Middle High German *ein glas bier-s* ‘a glass of beer’, Bauer 2017:33) in that the noun is not marked in a way that differs across enumerated and non-enumerated contexts. Sortal classifiers, by contrast, are not limited to uncountable nouns, and have been defined variously as:

- A member of a paradigm that forms a binary phrase with a numeral, which in turn forms a binary phrase with a counted noun (Lehmann 2000)
- A grammatical element that occurs with nouns (regardless of their degree of countability) in construction with numerals (Gil 2013)
- An expression that indicates a unit of counting or measure (Doetjes 2012)

Unlike mensural classifiers, sortal classifiers cannot be modified by an adjective: expressions like *ein kaltes Glas Bier* ‘a cold glass of beer’ have no equivalent with sortal classifiers; sortal classifiers generally cannot co-occur with mensural classifiers, e.g., Maithili *das(*-ṭā) kap cāy* ‘ten (*CLF) cup tea’ (Burghart 1992:I:117). A third type of numeral classifier designates

groups, similar to English *a flock of birds* (e.g., Beckwith 1998:131–133). Such classifiers indicate a set larger than one, including a pair. Sortal and mensural numeral classifiers do not indicate any number on their own, but are differentiated by the type of noun they occur with. Most languages exhibit numeral classifiers that refer to measures and groups. This study exclusively focuses on sortal numeral classifiers, which are much less common cross-linguistically and exhibit a very specific geographic distribution.

While a numeral classifier can provide an index to inherent semantic properties of the head noun (its use is often dependent on these properties) and can also carry pragmatic meaning, it does not achieve semantic modification in the way that adnominal elements such as adjectives do.¹ Furthermore, anaphoric use is common, in which case the head noun is excluded. We are additionally agnostic to the status of classifiers as a full phonological word, since they at times fuse phonologically with the numeral.

2.2 Optionality of plural marking

In a given language, individual nouns with plural reference may differ in terms of whether plural number must, cannot, or may be morphologically marked on them. Several Indo-Iranian languages, particularly older ones, have rigorous rules requiring that all semantically plural nouns take plural marking. In some Indo-Iranian languages, phonological and morphological change has resulted in paradigms where morphological plural cannot be marked on some noun types in some cases, i.e., where singular and plural forms are formally identical. In the remaining Indo-Iranian languages, plural marking on some noun types is optional, though it is rarely the case that optional plural marking is allowed on all noun types and pronouns; it is usually required on first person pronouns, at the very least, and tends to be sensitive to referential scales such as the animacy hierarchy (Silverstein 1976). Noun types which take optional plural marking in plural referential contexts are said to exhibit transnumerality or general number (cf. Corbett 2000:9–19), and it is generally the case that certain kind-denoting nouns can be partitioned into entities via strategies other than plural marking, rendering plural marking on such nouns as optional at best, if even allowed.

In the *World Atlas of Language Structures*, Haspelmath (2013) establishes three degrees of plural optionality (impossible, optional and obligatory), cross-classified against animacy types. In this coding scheme, a language is coded as having obligatory number marking when the distinction between singular and plural is neutralized in the context of numerals and other quantity words. Since here we are interested in the interaction between numerals, classifiers and number marking we opt for a coding scheme that keeps these three dimensions distinct. Specifically, we code any variation in nominal number marking as optional marking, regardless of whether there is a concomitant numeral (or numeral and classifier) in the same noun phrase. For present purposes we also gloss over distinctions in animacy contexts, or

¹In some languages, elements identical to numeral classifiers also occur in bare use with nouns, marking them variously for definiteness or indefiniteness (Simpson et al. 2011). Though these are often referred to as “bare numeral classifiers,” they fall outside the scope of our definition.

any other semantic or pragmatic dimension that might regulate the appearance of specific number markers.

2.3 The relationship between numeral classifiers and number marking

The best-known formulation of the observation that languages with numeral classifiers tend to have optional plural marking on at least some noun types (in constructions with numerals as well as those without them) comes from Greenberg (1972) and Sanches and Slobin (1973).² This generalization, known as the Greenberg-Sanches-Slobin (henceforth GSS) hypothesis (permutations of the names may vary from work to work), is borne out by a large number of languages. At the same time, there are several exceptions to the generalization (e.g., Aikhenvald 2000:100–101), making it clear that the generalization is not an absolute universal. Tang and Her (in press) find statistical evidence for this hypothesis using a large, diverse cross-linguistic sample.

The GSS hypothesis has clear counterexamples. For example, the Dravidian language Kurux has both plural marking and numeral classifiers. In this language, human nouns are obligatorily marked for plurality (Tirkey 2017:76). Therefore, a numeral classifier construction with a human noun also necessarily takes a plural marker.

(2) Kurux (Dravidian)

sa:t -j^han kuke -xadd -ar
seven -CLF girl -child -PL
'seven daughters? (Tirkey 2017:389, shortened)

Another counter-example has been noted for the Kiranti language Belhare, where both classifier use and number marking on human nouns is obligatory:

(3) Belhare (Kiranti)

sip -paŋ maʔi -chi
two CLF person NON.SING[ABS]
'two people (Bickel 2003:563)

Still, the GSS hypothesis appears to represent a dominant statistical tendency, and a number of proposals have been put forth to explain why plural marking is optional in many languages with numeral classifiers.

A prominent theory proposes that numeral classifiers help or are needed to enumerate, individuate or partition kind-denoting nouns, i.e., nouns like WATER or RICE that involve non-individuated and uncountable reference. The theory furthermore proposes that languages differ in their proportion of kind-denoting nouns (Quine 1960, Silverstein 1976, Lucy 1992,

²Cassirer (1923:190) mentioned that languages with numeral classifiers often do not have a distinction between singular and plural.

Croft 1994, Krifka 1995). The statistical version of the GSS generalization follows from these two proposals: languages with more kind-denoting nouns are expected to be more likely to use numeral classifiers in the service of enumeration; furthermore, since kind-denoting nouns are inherently uncountable, number marking is expected to be absent or at best optional on them. Most versions of this theory assume that the proportion of kind-denoting nouns is constrained by a referential scale, e.g., with human-denoting nouns being less likely to be kind-denoting than, say, food-denoting types (Lucy 1992). Theories differ, however, whether the variation involves material implications or merely lexical specificity. The material view argues that noun types differ cross-linguistically in their ontological properties: unlike count nouns, kind-denoting nouns designate masses and material without attention to shape and form, and this has ramifications for cognitive domains beyond language (Cassirer 1923, Lucy 1992, Imai and Gentner 1997). Under a lexical view, noun types vary cross-linguistically according to whether the distinction between kind and entity is specified in the lexicon or whether it is lexically ambiguous (Bisang 2002, 2017).

An alternative theory derives an absolute version of the Greenberg-Sanches-Slobin generalization from universal structural configurations. Thus, generative accounts hold that classifiers and plural markers occupy the same structural position (Borer 2005). This predicts the incompatibility of numeral classifiers and plural marking. When they do co-occur nevertheless, as in Kurux or Belhare, the relevant markers are predicted to differ from those in other languages, either because of different formal properties (e.g., the Belhare classifiers might not be real classifiers) or independent surface phenomena (e.g., the number marker appears where it does for phonological, not syntactic reasons).

However, other theories reject the Greenberg-Sanches-Slobin generalizations and derive the presence of numeral classifiers from properties that are not related to number marking: Aikhenvald and Dixon (1998) derive them from a general typological variable of reference classification. Gil (2013) sees them as an arbitrary conventionalization, possibly related to less configurational noun phrases (Gil 1987). Lehmann (2010) argues that in some languages they are simply necessary to give a numeral the status of a full word. Under these theories, the distributions of numeral classifiers and optional number marking reflect independent historical contingencies, especially effects of language contact.

In what follows we assess the Greenberg-Sanches-Slobin generalization empirically, probing the evidence for or against a diachronic correlation between numeral classifiers and optional number marking. While overall the Indo-Iranian data seem to be in line with the correlation, a number of observations cast doubt on it:

- Optional plural marking exists in a number of Indo-Iranian languages, likely a diachronic consequence of the loss of final syllable nuclei, and many of these show no sign of developing numeral classifiers. More generally, general number is not a strong predictor for the development of numeral classifiers in Indo-European. Some IE languages (e.g., Hittite) have optional plural on nouns with numerals only, and nowhere else; in some IE languages (e.g., Breton), singular number is even compulsory with numerals. None of these languages, however, developed classifiers.
- Modern Persian requires plural marking in certain referential contexts, in which case it

can co-occur with numeral classifiers, flying in the face of the apparent incompatibility of overt plural marking and classifier use.

- A not insignificant number of Indo-Iranian languages, such as Kumzari, Yaghnobi, Pashto, and Sinhala, have sortal numeral classifiers and obligatory plural marking

These observations raise serious questions as to whether the Indo-Iranian patterns owe to development of general number and a subsequent need to partition kind-denoting nouns. In view of this, we turn to statistical modeling to assess the hypothesis.

3 Enumeration in Indo-Iranian languages

Full information regarding the Indo-Iranian languages surveyed in this paper can be found in the Appendix. Here, we give a synopsis of the behavior seen across Indo-Iranian in the domain of enumeration, taking into account all attested chronological stages, highlighting examples which we believe to be important.

The presence of classifier systems in Indo-Iranian languages has attracted a fair amount of interest, particularly in contact linguistics (Thomason and Kaufman 1988:85ff.). Numeral classifiers are concentrated in the east of the Indo-Aryan-speaking region (with some possible exceptions described below); Assamese, the easternmost I-A language, has the largest number, with at least a dozen. The proximity of Ind-Aryan languages with numeral classifier systems to mainland Southeast Asia is conspicuous. Emeneau (1956, 1980) identifies numeral classifiers as a marker of the Indian linguistic area, but notes also that Indian numeral classifier systems “look like a western outlier of an area whose centre is in East and Southeast Asia” (Emeneau 1980:33). Matisoff (1978:78) states that “it seems obvious that the Nepali and Bengali classifier systems are due to T[ibeto]-B[urman] influence, while other Indic languages far removed from the TB area show no signs of developing classifiers.”

Heston (1980:147–148) shows that many features which serve as the basis for establishing India as a linguistic area, among them numeral classifiers, are found in Iranian languages as well. She additionally makes the following claim: “Lacking any contrary evidence, there seems no reason to assume the feature [i.e., numeral classifiers] is absent, rather than under-reported, in other Iranian languages [besides Persian and Pashto].” A survey of the evidence shows that the Central Iranian plateau is indeed a hotbed for formally diverse numeral classifier systems, whereas a handful of East Iranian languages appear to have borrowed classifiers from Dari, Tajik or other dialects of Modern Persian.

These views stand in contrast to Hackstein’s (2010) study on nominal classification among the Indo-European languages; for Hackstein, numeral classifiers found in Indo-Iranian languages are the grammaticalized endpoint of a family-wide tendency toward apposition between generic and non-generic nouns, though it remains unexplained why the distribution of numeral classifiers within Indo-European is so restricted.

3.1 Number marking

The Old Indo-Iranian languages Sanskrit, Avestan, and Old Persian have a rich morphological case and number system, and mark singular, dual, and plural number on nouns with the corresponding semantic number.³ Several Middle Indo-Iranian languages show this behavior as well. Pali, the Indo-Aryan language of the Theravada Buddhist Canon, generally maintains a clear morphological distinction between singular and plural; although the nominative singular and plural of *ā*-stems fell together due to regular sound change, a secondary plural suffix was recruited as number marking between the two numbers (Oberlies 2001:150–1). The Middle Iranian languages Khotanese Saka and Khwarezmian consistently mark plural number on nouns with plural reference. For Khotanese Saka, this is largely due to the preservation of a case and number system similar to that of Old Iranian; for Khwarezmian, this may be due to the fact that in a large subset of nouns, morphologized phonological processes such as palatalization render the singular stem distinct from the plural stem, e.g., *'kwnd* ‘finger’ vs. *'kwnc-n* ‘fingers’ (Durkin-Meisterernst 2009).

By contrast, a number of Middle Iranian languages do not consistently mark plural on nouns with plural reference, particularly for enumerated nouns. As seen in the following example, plural marking on Middle Persian nouns is entirely optional, particularly when the noun is modified by a numeral:

- (4) *ud čahar-dah dar ud mān panz ud gāh sē*
and 14 door and house 5 and throne 3
‘and fourteen doors and houses five and thrones three’ (Skjærvø 2009:223)

The same pattern holds for Parthian (Durkin-Meisterernst 2014:271). In late Bactrian, case and number distinctions have been neutralized due to the loss of distinctions between final vowels, resulting in an unmarked form without an ending which may be used with either singular or plural reference, and a marked plural form (Sims-Williams 2007:40). Some Sogdian heavy stem nouns show a form that is identical to the singular in plural contexts, e.g., *aβt paxarē-t* ‘seven planets (pl.)’ vs. *aβt paxarē* ‘seven planet (sg.)’ (Yoshida 2009:313). It is worth noting that Persian, Parthian, and Bactrian nouns tend, in contrast to those of Khwarezmian, to have plural forms that are a straightforwardly affixal extension of a singular “base” form, with no stem alternation (this is true as well for Sogdian heavy stems, in nominative case); this property has been associated with the presence of optional plural marking (Acquaviva 2004:352–4), as there is no overt element that marks singular nouns as unambiguously singular. Old Indo-Iranian languages, in contrast, tended to have plural marking involving more than simply adding an affix to the singular form, e.g., Sanskrit *dev-a-ḥ* ‘god (nom.sg.)’ vs. *dev-ā-ḥ* ‘god (nom.pl.)’. In these languages, historical phonological and morphological changes affecting final syllables often resulted in formally identical singular and plural forms, with optional plural marking carried out by suffixes that were previously collective (e.g., Middle Persian *-ān*, *-hā*; Sogdian *-t*), or somehow yielded a similar extensional pattern.

³Isolated Vedic forms show singular number on nouns with plural reference (Oldenberg 1909).

Modern Indo-Iranian languages show several different patterns. For some languages, plural marking is obligatory. Certain languages of northern Pakistan such as Palula, Kalam Kohistani, and Dumaki consistently mark plural number on nouns with plural reference. Number marking is obligatory in Sinhala, which contains a complex and opaque system comprising at least three noun types: those where the singular and plural are derived from a common base, those where the plural is derived from an unmarked singular, and those where the singular is derived from an unmarked plural. This system appears to have come about via a complex series of developments: initially, plural suffixes were lost in some nouns, leading to a state of affairs where singular marking was optional (Nitz and Nordhoff 2010). In non-enumerative contexts, Ossetic consistently marks plural nouns with the suffix *-t-*, cognate to the Sogdian plural suffix; when a noun is enumerated by a numeral greater than one, the noun is marked by the suffix *-i* (Digor)/*i* (Iron), synchronically identical to the genitive suffix. According to Thordarson (2009:132), this suffix continues the Old Iranian plural suffix **-ah*. The author links this diachronic behavior to that of Yaghnobi, where nouns are marked for oblique case suffix *-i* (perhaps < **-ah*) when enumerated.

Optional plural marking is found in a large number of contemporary languages, including Modern Persian, Kurdish, Zazaki, Bengali, Maithili, Dhivehi, and others. In line with global expectations, the optionality of plural marking in many of these languages is dependent on referential scales, with plural marking often required on nouns of higher animacy. No Indo-Iranian language in our sample allows optional plural marking on pronouns, although some third person pronouns have no morphological distinction between singular and plural.

Another pattern, widespread in Modern Indo-Aryan, involves noun paradigms with morphological restrictions on the expressibility of plural number. For Hindi consonant-final masculine nouns, the direct singular is formally identical to the direct plural, and distinctions in number can be overtly realized only in oblique case forms. Near-identical restrictions of this sort are found in Panjabi, Sindhi, and adjacent Indo-Aryan languages. Similar morphological restrictions can be found in isolated Iranian languages. In Rakhshani Balochi, nouns can be marked for indefiniteness and singularity via the suffix *-e*, but otherwise, there is no morphological distinction between singular and plural (Barker 1969:3ff.). In Sangesari, plural is consistently marked on oblique nouns, but cannot be marked on direct nouns, except for a restricted set of items (Azami and Windfuhr 1972:70ff.). Space does not permit a full investigation into the forces responsible for the development of optional plural marking in Indo-Iranian, though this will undoubtedly prove to be a valuable research direction.

3.2 Classifiers

The diachrony of Indo-Iranian numeral classifier systems is not well understood given lacunae in the historical record, making it difficult to determine whether they developed as a response to the loosening of restrictions on plural marking, because of contact, or due to other factors. Attestations of classifiers are largely absent from pre-modern Indo-Aryan languages, e.g., Old Bengali, perhaps due to stigmatization and suppression in the literary register (for discussion, see Barz and Diller 1985:168). It is however possible to trace the

development of certain numeral classifiers in the history of Persian, though the material is incomplete and the exact pathway of development is somewhat ambiguous. In the absence of evidence from this historical record, it is in theory possible to infer whether classifiers developed due to language contact on the basis of (1) the presence of matter borrowing and (2) languages' proximity to groups with numeral classifiers, but again, the picture is not entirely clear.

On the whole, Indo-Iranian classifiers are a mix of inherited and borrowed material. Agia Varvara Romani has borrowed the classifier *-tane* from Turkish⁴ (and has obligatory plural marking):

- (5) *Dikhlém pándžtane raklá*
 see.1SG.PST 5.CLF girl.PL.ACC
 'I saw five girls' (Igla 1996:45)

The classifiers found in Indo-Aryan languages of South Asia tend to belong to a core group of elements with transparent Old Indo-Aryan (OIA) etymologies, which are supplemented with additional classifiers (Assamese has roughly a dozen numeral classifiers, all of which appear to be from inherited Indo-Aryan material). The most geographically widespread Indo-Aryan classifier is a reflex of Old Indo-Aryan *jana-* 'person', occurring in Sinhala as the element *denaa* (Geiger 1942:4) as well as in Eastern Indo-Aryan languages. Another common classifier (Nepali *vaṭā*; Bengali, Oriya, *ṭā*; Assamese *tā*) is derived by Chatterji (1926:684ff.) from OIA *ṛt-ti-ka-*, a deverbal noun built to the root *vart-* 'turn'. The classifier *goṭ*, found in Maithili and other Eastern Indo-Aryan languages, continues Old Indo-Aryan *gōṭṭa-* *'something round' (Turner 1966:229). With the exception of Sinhala, the aforementioned languages have optional plural marking on most noun types, and tend to prohibit plural marking on enumerated nouns that co-occur with numeral classifiers:

- (6) Maithili
sāt -ṭā murgī (-sabh)*
 seven CLF hen PL
 'seven hens' (Burghart 1992:117)

However, in Nepali, referential scales may require plural marking in some circumstances, leading to the co-occurrence of numeral classifiers and plural marking:

- (7) Nepali
*cār janā mitra *(-haru)*
 4 CLF friend PL
 'four friends' (Bhim Lal Gautam, p.c.)

⁴The Turkish form is generally thought to be an Iranian loan cognate to Persian *dānah* 'grain', with devoicing of initial *d-* (Stilo 2018:138).

Iranian languages employ a more varied mix of inherited forms and borrowed elements from Arabic as well as Turkic languages. Modern Persian employs a number of Arabic terms in addition to the inherited classifier *tā*. It seems unlikely that Persian borrowed these items as classifiers *per se*, since the syntactic patterns characterizing Persian classifier use differ from those of Arabic dialects with numeral classifiers.⁵ These classifiers of Arabic origin can be found in related Iranian languages. Classifiers of Turkic origin are found as well. Zazaki *teney* is a Turkish loan (the same element is found in Agia Varvara Romani), and Pashto *tana/teni* may be from a Turkic source as well, though these forms are ultimately Iranian back loans. The Sariqoli classifier *tol* may be of Turkic origin (cf. Uygur *tal*).

The full range of circumstances under which Iranian classifiers developed is unknown, but Middle and Early Modern Persian provide a window onto the usage of the precursor of the Modern Persian general classifier *tā*, which continues Middle Persian (Pahlavi) *tāg*, a multifunctional element glossed as ‘item, unit, alone, single’ by MacKenzie (1971), who keeps this headword separate from *tāg* ‘branch’. The following examples from the Pahlavi Widēwdād show a diverse range of uses:

- (8) *nay ēw tāg*
reed one piece
‘a single piece of reed’ (Moazami 2014:272)
- (9) *ēw tāg frāz ō ātaxš dahēd*
one piece forward to fire give.IMPER
‘give once to the fire’ (Moazami 2014:254)
- (10) *bōb -ē bālišn se tāg*
fine_carpet INDEF pillow three piece
‘a fine carpet (and) three-fold cushions’ (Moazami 2014:352)
- (11) *spiš -ē ayāb rišk -ē tāg*
louse EZ or nit EZ piece
‘a single louse or nit’ (Moazami 2014:390)
- (12) *u -š nō -īh tāg wēd barīd*
he 3SG nine ABSTRACT.SUFFIX branch willow bring.PST.3SG
‘he brought nine branches⁶ (of barsom)’ (Moazami 2014:468)

Mache (2012:171) cites the following example, in which *tā(g)* appears to be used as a sortal classifier:

⁵For instance, Persian employs a numeral classifier *ra*’s ‘head’ (< Arabic) for animals. Greenberg (1972:18–20) cites constructions from a 19th century Arabic dialect of Oman and Zanzibar which employs the same word when counting animals, but in contrast to Persian usage, it inflects for number.

⁶MacKenzie (1971) keeps the headwords for ‘branch’ and ‘item’ separate. Regardless of whether these entries should be separate, the use of ‘branch’ as a mensural classifier here is noteworthy. We see also the reduplicative plural *tāg tāg* (Moazami 2014:122).

- (13) *čand tā dānāgān ī hindūgān*
 some tā wise.PL EZ Indian.PL
 ‘some wise Indian men’

She argues on the basis of this example that Middle Persian is a classifier language, though the evidence is quite restricted. We refrain from treating Middle Persian as a numeral classifier language, given the scant, rather ambiguous evidence and the fact that the loose morphosyntactic integration of information in noun phrases containing *tāg* makes it difficult to determine whether they meet the criteria we have chosen for numeral classifiers.

Numeral classifier constructions like those seen in Modern Persian are not attested in Early Modern Persian. However, some constructions involving *tā* ‘piece, unit’ are found in Early Modern Persian. In these constructions, *tā* marked for indefiniteness is followed by a number or quantifier.⁷ Most instances are limited to one text, the Iskandar-Nāmah, which can be dated to the 12th century CE (Lazard 1963:127), but this usage is conceivably reflective of non-literary usage (Lazard 1963:217–218):

- (14) *Zangiyān i nīmkušta tā’ē čand*
 Zangi.PL EZ half.dead piece.INDEF few
 ‘some half-dead Zangis (pej. ethnic term)’
- (15) *pariyān rā tā’ē saδ bā rasūl bifirist*
 Peri.PL OBJ piece.INDEF hundred with messenger send.IMPER
 ‘Send about a hundred Peris, with the messenger’
- (16) *tā’ē duvēst rā az Zangiyān bikuštand*
 piece.INDEF two.hundred OBJ of Zangi.PL kill.PST.3PL
 ‘they killed two hundred Zangis’

In the above examples, the appositional, loose morphosyntactic integration of elements into the noun phrase is striking, as well as the discontinuity of the noun phrase. In one example, the numeral element is followed by the object marker *rā*, in another, the noun that the numeral element modifies. Adverbial use of *tā’ē* constructions is found as well:

- (17) *tā’ē čand bar ān zan zaδ*
 piece.INDEF few DAT this woman hit.PST.3SG
 ‘he dealt some blows to this woman’

Additionally, it is worth noting that in the above examples (though they are few in number), *tā’ē* constructions cooccur with nouns with overt plural marking, while prenominal numerals cooccur with unmarked nouns. The exact circumstances under which Early Modern Persian constructions came to evolve into ModP numeral classifier constructions remain unclear, if the Early Modern Persian pattern is in fact the diachronic precursor of the ModP one.

⁷Lazard states that this construction — e.g., indefinite nouns followed by a number or quantifier — indicates an order of magnitude for large numbers (“nombres ronds”), and is used for approximation with small numbers.

However, among its multifunctional uses, it is apparent that *tā* serves as an optional means of integrating numeral elements into the noun phrase at both diachronic stages, albeit with differences in the order of the numeral and *tā*, as well as differences in rigidity of the placement of the numeral element with respect to the noun being modified.

It is clear that classifier-like uses of *tāg* were on the rise during Middle Persian times. This is roughly the earliest date at which Turkic and Iranian languages were in contact (Golden 2006). It is possible that Turkic influence led to the conventionalization in Early Modern Persian of this incipient tendency toward classifier-like constructions, though we have no overt evidence of Turkish influence in the form of matter borrowing (conversely, several Turkic classifiers are made up of borrowed Iranian matter).

Although the Persian historical record provides a slender window onto their development, the conditions that gave rise to numeral classifiers across Indo-Iranian are not entirely clear from the empirical coverage available. We hope to shed further light on the origins of Indo-Iranian numeral classifiers using a probabilistic methodology capable of quantifying the most likely trajectories of classifier development in this subgroup.

4 Data and Methods

We employ an explicitly phylogenetic method to address the GSS hypothesis. Given a phylogenetic representation of the languages in our sample, and assuming that change in the linguistic features studied in this paper can be modeled according to a continuous-time Markov (CTM) process, we can quantify and approximate the temporal rates at which transitions between different feature variants occur. These rates can be used to operationalize a wide range of questions regarding the diachronic dynamics of the features under study, and can be used to reconstruct probable trajectories of their development.

4.1 Data

The phylogenetic comparative methodology that we employ involves two key ingredients: a tree sample of 1000 phylogenies of the languages under investigation, and a featural representation of each language of interest.

Our tree sample consists of a hand-crafted topology (i.e., branching structure), with branch lengths drawn stochastically in order to simulate phylogenetic uncertainty. We combined insights regarding higher-order subgroups within Indo-Aryan and Iranian from the comparative-historical literature (Oranskij 1977, Masica 1991, Toulmin 2009, Wendtland 2009) with information from work in progress in ongoing phylogenetic inference projects involving lexical and phonological characters which allowed us to resolve questions regarding shallower subgroups in the Indo-Iranian tree. Chronology is largely uncontroversial within Indo-Iranian, and the basal patterns of our tree sample are in line with the Indo-Iranian chronology outlined by Kuz'mina (2007).

The data consist of feature codings for each language according to two variables of interest, concerning (1) the presence of numeral classifiers and (2) optionality of and

restrictions on plural marking. These variables show the following attested values:

1. Numeral classifiers

- (a) Classifiers present: +CLF
- (b) Classifiers absent: −CLF

2. Plural marking

- (a) Optional plural marking in at least some contexts: OPT.PL
- (b) Morphological restrictions on plural number: MORPH.PL
- (c) Obligatory plural marking in all contexts: OBL.PL

Combinations of values for these variables in our sample can be seen in Figure 2. To ensure that our inference procedure is tractable and meaningful, we keep the number of levels for each variable low. We treat languages with morphological restrictions on plural marking as having obligatory plural marking, since languages with restricted plural marking tend to mark plural number on nouns to the extent possible, whereas languages with optional plural marking choose not to mark plural number in contexts where it is possible. This leaves us with a single binary feature $\pm\text{OBL.PL}$. Furthermore, we treat all four attested combinations of values of $\pm\text{CLF}$ and $\pm\text{OBL.PL}$ as a single feature with multiple values; e.g., Bengali has the value (+CLF, −OBL.PL).

4.2 Model and Inference

CTM models of character evolution We model changes between different feature states via a common phylogenetic comparative method, the CTM process of character (i.e., feature) evolution. Under such a model, transitions between different states (i.e., feature variants) take place at non-negative evolutionary **RATES**, the inverse of which represents the average time the system spends in a given state. Rates between different states can be found in the off-diagonal cells of the instantaneous rate matrix Q ; diagonal cells of the matrix take values such that rows sum to zero. For a given timespan t , the row-stochastic matrix P_t of transition **PROBABILITIES** between all states (along with self-transitions) can be computed via matrix exponentiation:

$$P_t = \exp \{Qt\}$$

We place prior distributions over the rates in Q such that transitions occur over realistic time intervals, and infer posterior distributions of each rate, as defined below:

$$P(Q|D, \mathbf{T}) \propto P(D, \mathbf{T}, Q) = \sum_{T \in \mathbf{T}} P(D, Q|T)P(T|\mathbf{T}) \approx \frac{1}{|\mathbf{T}|} \sum_{T \in \mathbf{T}} P(D|T, Q)P(Q)$$

D represents the observed linguistic data; \mathbf{T} represents the sample of trees. The probability of the data given a tree and set of rates, $P(D|T, Q)$, can be efficiently computed via the

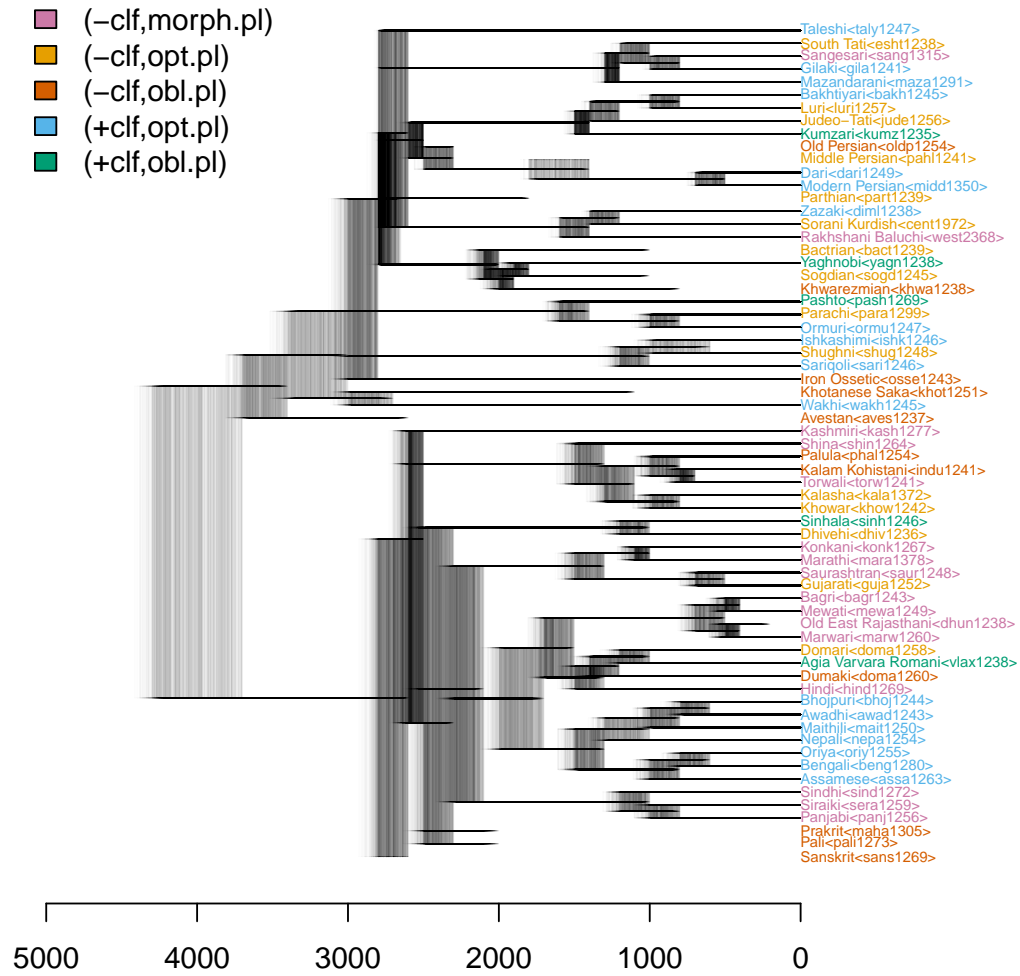


Figure 1: Sample of 1000 Indo-Iranian phylogenetic trees; tip colors represent languages' states

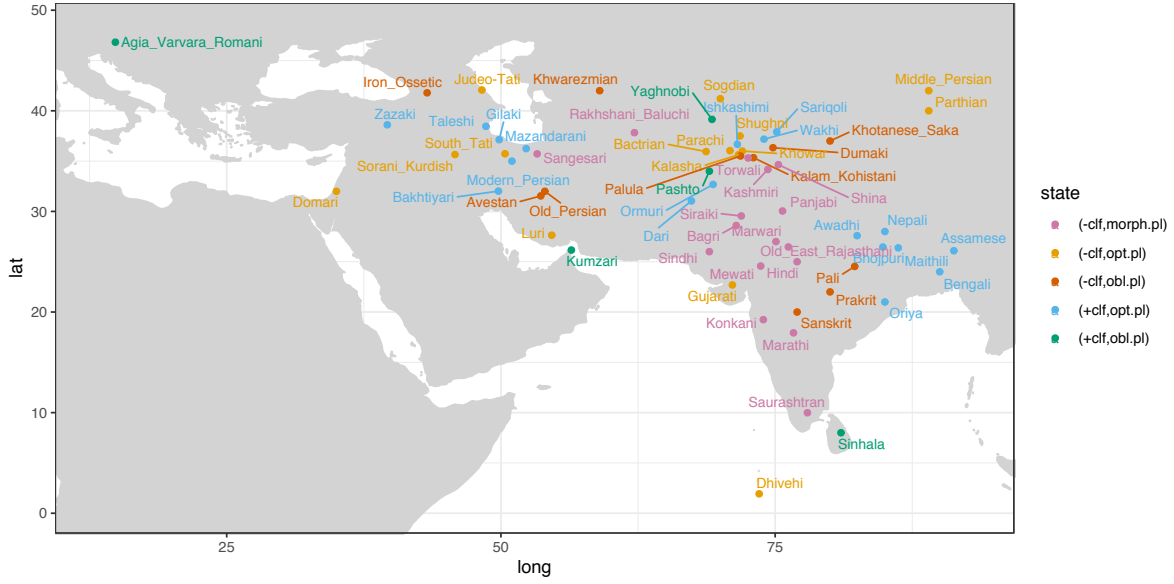


Figure 2: Approximate locations of languages in sample, based on closest glottocode matches; \pm CLF stands for presence/absence of sortal numeral classifiers, \pm OBL.PL for presence/absence of obligatory plural on all nouns, and MORPH.PL for morphologically restricted obligatory plural

Pruning Algorithm (Felsenstein 2004:251–5). Once the posterior distributions of the rates are inferred, they can be used to reconstruct the probability of a given character state at internal nodes of the tree (i.e., nodes where no data are observed). Posterior rates can also be used to carry out stochastic character mapping (SCM; Nielsen 2002, Huelsenbeck et al. 2003, Bollback 2006), an iterative process which samples locations on branches of the phylogeny where changes between states have the highest posterior probability of occurring.

Phylogenetic hypothesis testing The literature on the relationship between classifier presence and optionality of plural marking surveyed above makes the prediction that certain pathways of diachronic development will be considerably constrained, if not impossible. Changes of the type $(-CLF, -OBL.PL) \rightarrow (+CLF, -OBL.PL)$ will be favored, while $(-CLF, +OBL.PL) \rightarrow (+CLF, +OBL.PL)$ will be infrequent to nonexistent.

In linguistics, phylogenetic comparative methods provide a means of testing for associations between pairs of linguistic features while controlling for phylogenetic relatedness among languages in the sample. A standard way of testing for correlated evolution between two discrete binary features, such as \pm CLF and \pm OBL.PL is Pagel’s (1994) DISCRETE model, which assesses the relative model fit of a dependent model, which constrains evolutionary rates in a manner thought to be compatible with correlated patterns of evolution, against a null, independent model, which models two independent character histories for each of the

features in question (Pagel and Meade 2006, Dunn et al. 2011). In the Bayesian context, a common practice is to carry out this assessment using Bayes Factors (i.e., the ratio of marginal likelihoods for each model).

We depart from the DISCRETE model and hypothesis testing with Bayes Factors due to several reasons. First, DISCRETE model has been shown to exhibit problematic behavior under certain circumstances (Maddison and Fitzjohn 2014); in particular, scenarios in which features undergo relatively infrequent changes over the tree can be prone to false detection of the presence of correlated evolution, though this is not a problem for all datasets.

Bayes Factors have traditionally been viewed as a lean way of comparing nested models, but statistical science is gradually moving away from their use in favor of alternative approaches. The reasons for this change are both technical and philosophical (Gelman and Shalizi 2013). Our key objections to using DISCRETE (along with Bayes Factors) are as follows:

- The DISCRETE model tells us whether there is support for interdependent evolution, but suppresses most of the dynamics of change over the tree, including directionality of change. Since directionality is built into our hypothesis, we prefer to observe rates from a single model in order to determine whether the classifiers develop more frequently in the presence of optional plural marking — not simply whether a change in one feature is followed by a change in the other feature.
- Bayes factors require the operationalization of the null and alternative hypothesis in terms of statistical models, and this may involve some degree of misspecification; this makes model comparison problematic, as misspecification may manifest itself in unpredictable ways which lead to erroneous results.

Given these concerns, we choose the computationally simpler approach of carrying out hypothesis testing within a single model (cf. Kruschke 2011) and allowing the more plausible evolutionary story to fall out of these results. Our model involves transition rates between all possible combinations of (\pm CLF, \pm OBL.PL), including transitions involving two state changes, e.g., ($-$ CLF, $-$ OBL.PL) \rightarrow ($+$ CLF, $+$ OBL.PL). Changes of this sort are not explicitly represented under the DISCRETE model, perhaps because such non-cascading changes are thought to occur with infinitesimal probability in comparative biology. However, we cannot be sure that developments in which multiple linguistic features undergo simultaneous intergenerational change are altogether nonexistent, or more specifically, that a change ($-$ CLF, $-$ OBL.PL) \rightarrow ($+$ CLF, $+$ OBL.PL) is impossible. In the case of intense language contact, speakers undergoing language shift often impose characteristics of their native language onto the language that they are in the process of adopting. The transfer of one language's profile to another is thought to be a gradual, incremental process (Ross 1996, 2007); at the same time, it is not inconceivable that speakers could simultaneously introduce multiple features from their native language into their second language. The validity of this assumption aside, if scenarios involving simultaneous change of two morphosyntactic features are impossible, it is preferable to allow this behavior to fall out of a statistical model's performance, rather than constrain the behavior of the model.

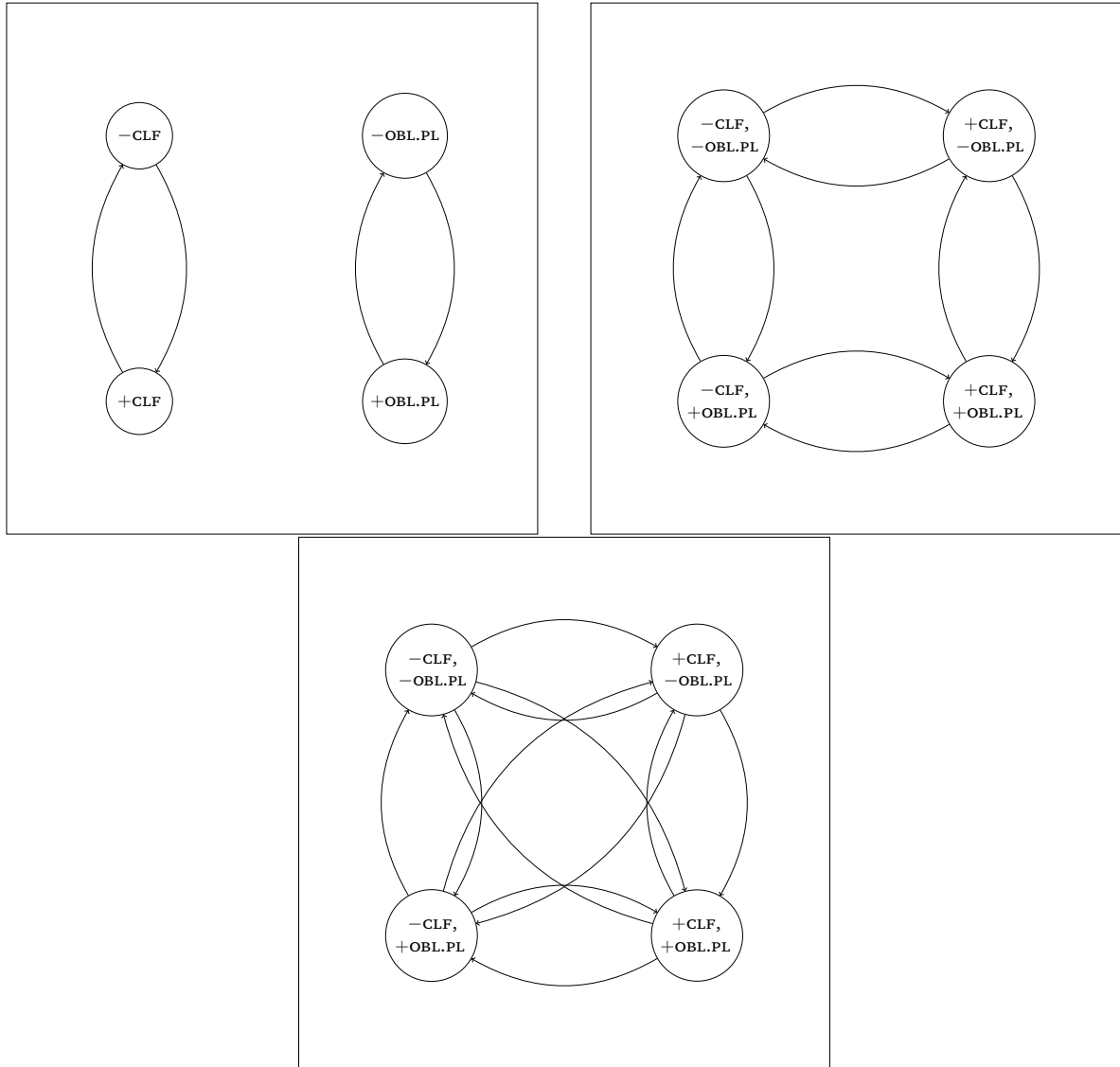


Figure 3: Top: independent (left) and dependent (right) models compared under the DISCRETE approach (Pagel 1994). Bottom: the single model used in our analysis. Note that our model explicitly allows transitions such as $(-CLF, -OBL.PL) \rightarrow (+CLF, +OBL.PL)$, and that the single model that we use has the same number of parameters as the two models compared under the DISCRETE model.

Inference We scale branch lengths by dividing them by 1000, and place a $\text{Gamma}(1, 1)$ prior over the rates Q , representing an average of one change per millennium. We infer posterior distributions for the evolutionary rates using the No U-turn Sampler of `RStan` (Carpenter et al. 2017), aggregating posterior values over trees in the sample. Posterior samples of evolutionary rates can be used to simulate ancestral states at each unobserved node and simulate character histories over the tree, shedding light on likely evolutionary trajectories involving the features of interest.

5 Results

5.1 Relationship between optional plural marking and classifier development

The posterior rates can be seen in Figure 4. The top three most frequent changes involve transitions away from the state (+CLF, +OBL.PL), showing that the combination of classifiers with obligatory plural is diachronically unstable. This is followed by transitions away from the state (−CLF, −OBL.PL), suggesting that lacking both classifiers and obligatory plural is a bit more stable, but still somewhat less stable than the other combinations.

With an eye to addressing the question of whether classifier development is mediated by optional plural marking, we further inspect larger-scale evolutionary dynamics in our data set by following a property of Continuous-Time Markov chains whereby a transition rate $q(A \rightarrow B \cup C) = q(A \rightarrow B) + q(A \rightarrow C)$ (Padma and Vijayalakshmi 2011:20). Concretely, we estimate the rate of the development of numeral classifiers in the presence of optional plural marking by summing over the rates for all transitions where the initial state has the value (−CLF, −OBL.PL) and the final state has the value (+CLF, ±OBL.PL), contrasted with the rate of classifier development in the presence of optional plural marking:

$$q(-\text{CLF} \rightarrow +\text{CLF} \mid -\text{OBL.PL}) = q((-\text{CLF}, -\text{OBL.PL}) \rightarrow (+\text{CLF}, -\text{OBL.PL})) \\ + q((-\text{CLF}, -\text{OBL.PL}) \rightarrow (+\text{CLF}, +\text{OBL.PL}))$$

$$q(-\text{CLF} \rightarrow +\text{CLF} \mid +\text{OBL.PL}) = q((-\text{CLF}, +\text{OBL.PL}) \rightarrow (+\text{CLF}, -\text{OBL.PL})) \\ + q((-\text{CLF}, +\text{OBL.PL}) \rightarrow (+\text{CLF}, +\text{OBL.PL}))$$

Figure 5 shows that the median rate of classifier gain given optional plural marking is higher than that given obligatory plural marking. Figure shows a density plot of the differences between the rate of classifier gain given optional plural marking and that given obligatory plural marking for all posterior samples. Values greater than zero indicate that classifiers are gained more frequently in the presence of optional plural marking than in the presence of obligatory plural marking, and is represented by the blue region under the curve. The fact that the posterior distribution over the differences overlaps with zero means that we

cannot reject the null hypothesis (i.e., that classifier development is not more frequent in the presence of optional plural marking) outright. However, the evidence in favor of the GSS hypothesis is considerably stronger than the evidence in favor of the null hypothesis. For 90.9% of posterior samples, the difference in rates is greater than zero, whereas the difference is less than or equal to zero in only 9% of samples. We interpret this as partial evidence for the GSS hypothesis, but it is clear that other factors must be at play as well. We investigate branch-specific developments in detail by carrying out stochastic character mapping, which allows us draw inferences regarding the most probable trajectory leading to the development of classifiers on each branch where they emerge in the tree.

We carry out stochastic character mapping using the SIMMAP method (Bollback 2006) as implemented in the R package `Phytools` (Revell 2012). We simulate character histories on the tree over 1000 iterations, drawing from the posterior sample of transition rates. The standard way for visualizing the aggregation of these histories is to use a density map, which represents the probability of a state in continuous space over the tree using a color gradient. Visualization can be a challenge for more than two states, since colors can become muddy in regions where uncertainty over the state value of the character is high. For this reason, we estimate a maximum a posteriori (MAP) character history over the tree by tabulating for each branch the counts for each type of transition history (ignoring the actual waiting times between transitions) and taking the most frequent transition history. We give both a density map and a MAP character history in Figures 6–7. Each branch is annotated according to the posterior support for its MAP transition history. This procedure also allows us to estimate the probability of that a given state is present at each internal node of the tree, i.e., nodes in our tree where data are not observed

We see that the most probable state for the oldest nodes in the tree is $(-CLF, +OBL.PL)$, which accords with received wisdom regarding prehistoric stages of Indo-Iranian. The states $(+CLF, -OBL.PL)$ and $(+CLF, +OBL.PL)$ emerge at various points on the tree. Changes to $(+CLF, +OBL.PL)$ tend to be quite recent, taking place on branches leading to individual languages. The state $(+CLF, -OBL.PL)$ is reconstructed with high posterior probability at several deep chronological nodes, notably the node ancestral to Modern Persian and Dari (with probability 0.7816), the node ancestral to the Pamir languages (0.5122), and a node ancestral to several Eastern Indo-Aryan languages (0.6017). In the following section, we analyze these developments individually, assessing possible contact events that may have led to the development of numeral classifiers in Indo-Iranian speech varieties.

6 Discussion

Overall, our results show that the GGS may account for some incidences of classifier development in Indo-Iranian, but suggest that other factors such as contact may play a role as well and that there may be some degree of interaction between these factors. In the following section, we integrate our model results with qualitative insights regarding Indo-Iranian contact history in order to sketch a plausible account of the development of classifiers across the subgroup.

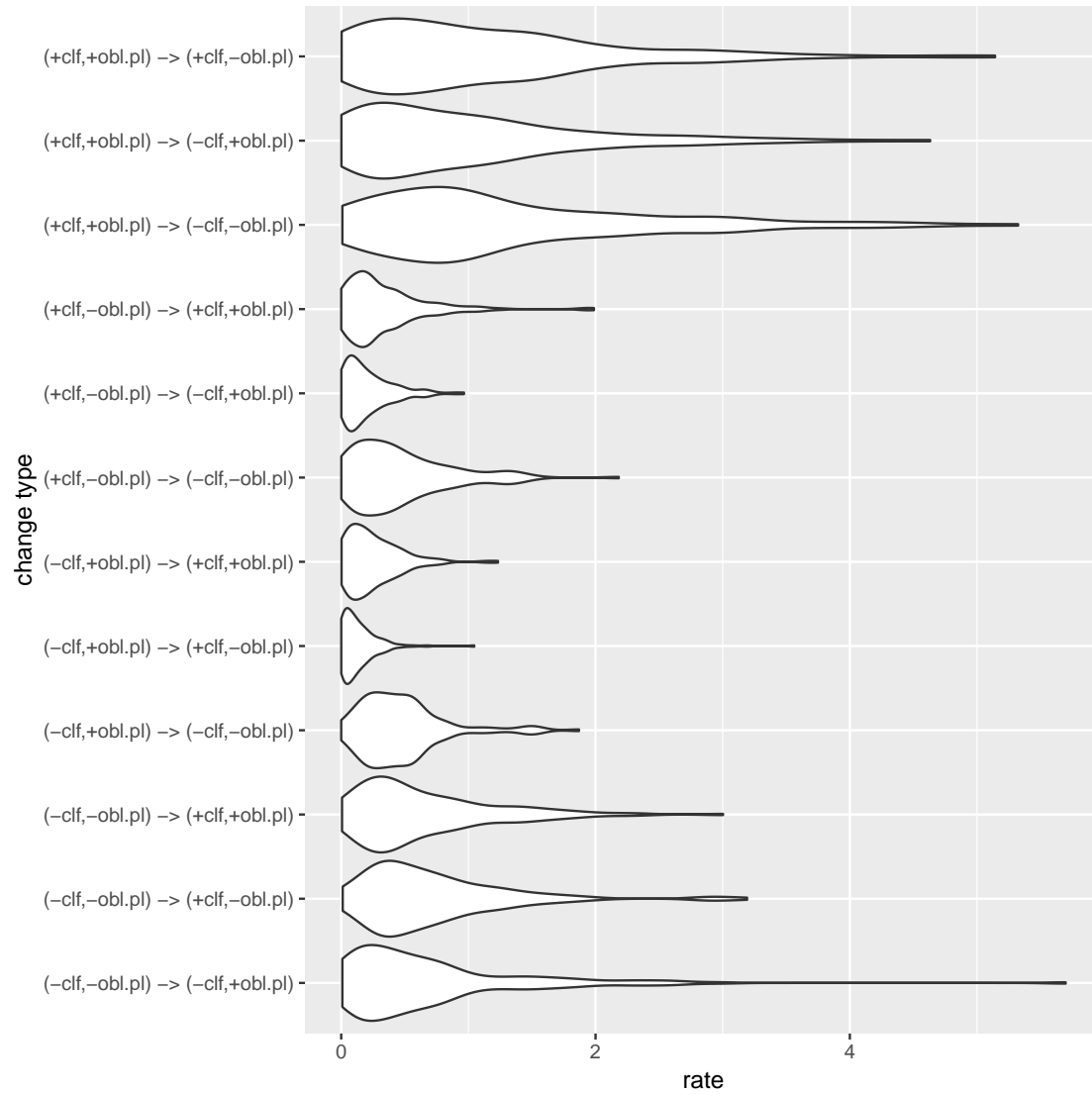


Figure 4: Posterior distributions of rates for each change type, with medians indicated. Abbreviations as in Figure 2.

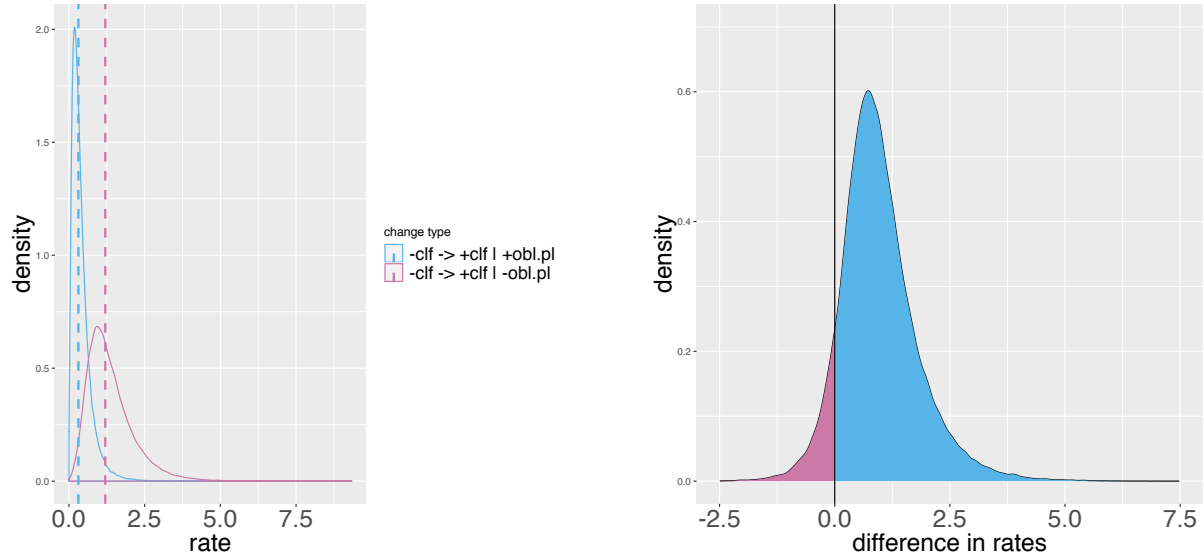


Figure 5: Left: posterior distributions of rates for classifier gain given obligatory plural marking versus classifier gain given optional plural marking. Right: Difference between rates of classifier gain in the presence of optional versus obligatory plural marking; values greater than zero indicate that classifiers are gained more frequently in the presence of optional plural marking as opposed to obligatory plural marking.

A striking aspect of the character history that we have inferred over the tree is that in Indo-Aryan languages, classifiers come about far less frequently than they do in Iranian. In each of the three cases of emergence, classifier development does not appear to be preceded by a period of optional plural marking, counter to the GSS hypothesis. This state of affairs seems to favor language contact as a potential explanation; however, matter borrowing has taken place in only one Indo-Aryan language with classifiers: Ajia Varvara Romani has clearly borrowed *tane* from Turkish (see above); this element is found in other Romani dialects in contact with Turkish (Matras 2002:204).

Reasons for the development of a numeral classifier in Sinhala are less clear, given the absence of borrowed matter pointing explicitly to contact; however, the Sinhala classifier construction is very similar to that of surrounding Dravidian languages, especially Tamil. Sinhala *denaa* apparently still has its meaning ‘people’ in other contexts (Chandralal 2010:passim). In Modern Literary Tamil (Lehmann 1993:112–114), numerals above one have a so-called pronominalized form with a suffix *-ar*. In Modern Spoken Tamil (Schiffmann 1999:132–135), pronominalized numerals higher than one add the noun *peeru* ‘name’ instead, which apparently can also mean ‘person’. In both languages the newly formed numeral can be used attributively or pronominally. If used attributively, it can precede or follow the head noun. In Literary Tamil, they have a marked genitive reading instead if preposed. In Spoken Tamil, they have a specific or definite reading if postposed. The similarities with Sinhala are striking.

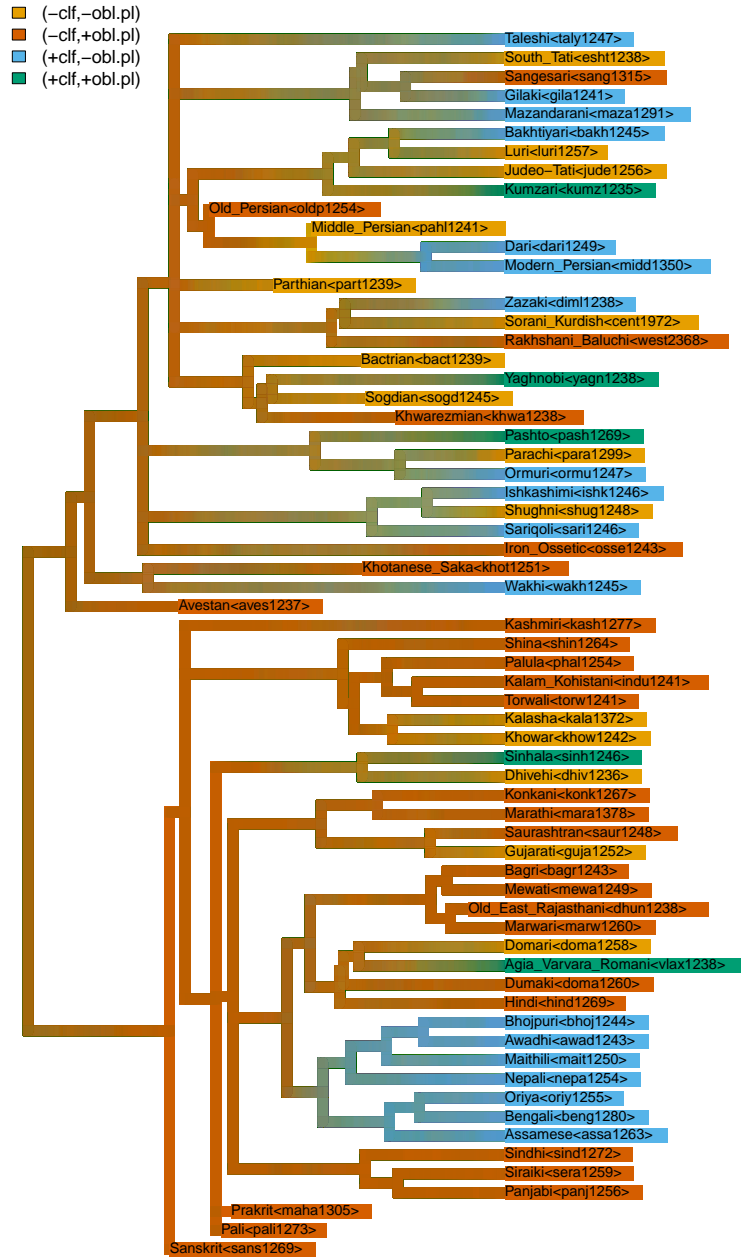


Figure 6: Density map aggregating probable character histories over maximum clade credibility tree constructed from the tree sample.

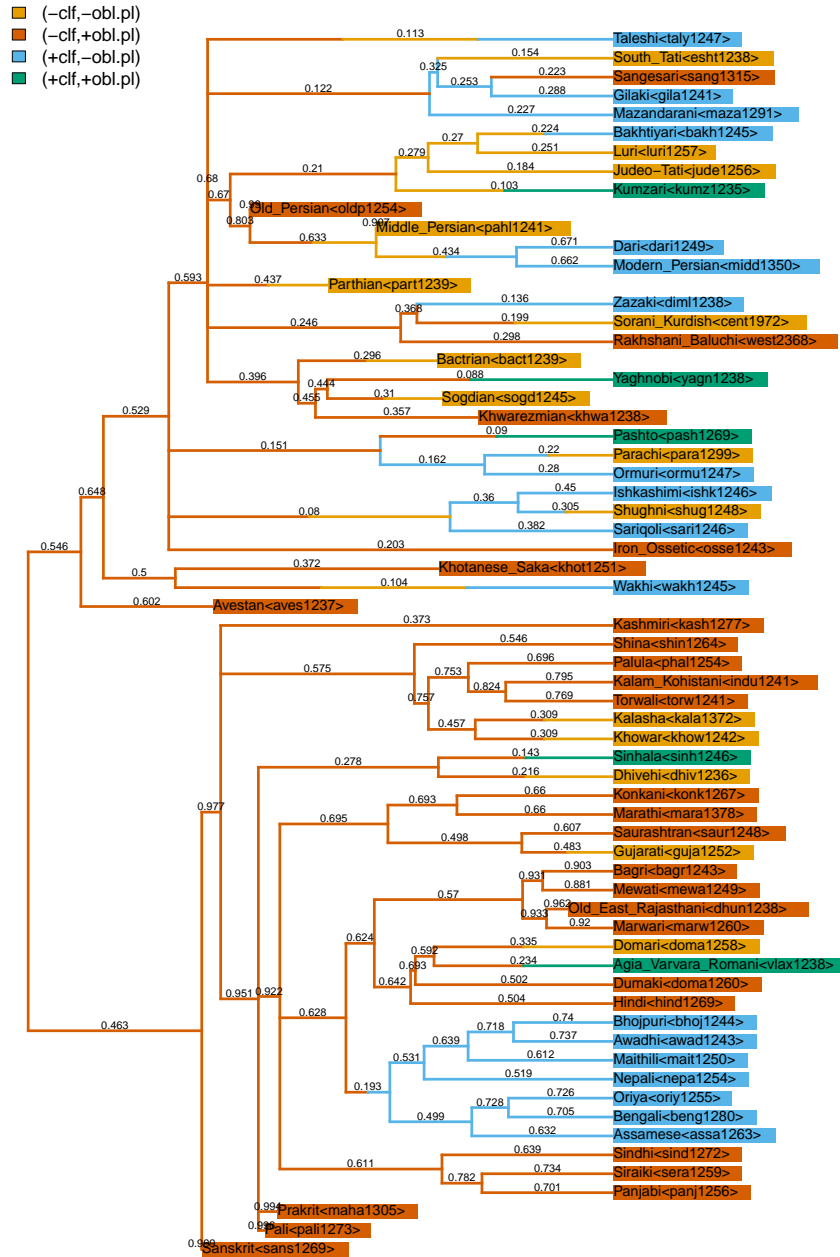


Figure 7: MAP character history over maximum clade credibility tree constructed from the tree sample.

In all three languages there is [N [Num Clf]] word order as at least one possibility and there is a connection with animacy, non-animate or non-human entities being unmarked. Some Dravidian languages on the mainland have similar classifier-like constructions for humans. In Telugu, for instance, numerals above eight combine with the word *mandi* ‘persons’, e.g. *padi-mandi* ‘ten persons’ (Krishnamurti and Gwynn 1985:106–109). However, the direction of influence is not altogether clear. More historical and diachronic research is necessary to clarify this point.

The diachrony of the development of numeral classifiers in Eastern Indo-Aryan languages like Bengali is poorly understood. There is some evidence that medieval East Indo-Aryan languages had optional plural marking. Mukherjee (1963:23) states that Old Bengali lacks morphological number, but that “plurality ... can be expressed by combining a singular noun with a number of words which denote plurality.” It is not clear whether this feature co-existed with classifier use, as there is good reason to believe that this feature was suppressed in literary registers, our only source of data on languages of this sort (cf. Barz and Diller 1985). If the historical scenario suggested by our results is accurate, a sweeping change of the type (–CLF, +OBL.PL) → (+CLF, –OBL.PL) took place in the history of Eastern Indo-Aryan. This development suggests a change in the general typological profile of these languages, possibly brought about by widespread language shift.

Eastern Indo-Aryan languages with numeral classifiers are spoken in the vicinity of Dravidian (e.g., Kurux), Austroasiatic (especially Munda, e.g. Kharia), and Sino-Tibetan (especially Tibeto-Burman, e.g., Jero) languages, many of which also exhibit numeral classifiers. Striking parallels with the behavior of East Indo-Aryan classifiers can be found in the Tai-Kadai language Khamti. In Khamti, [N Clf Num] word order has an indefinite reading with the numeral ‘one’, while [N Num Clf] word order is definite. Interestingly, Assamese, Bengali, and Oriya have a connection of word order with definiteness, where [N Num Clf] word order is definite while [Num Clf N] order is indefinite.

(18) Khamti (Tai-Kadai)

*kuun*⁴*maau* *koo*¹ *leeung*³
 bachelor CLF one
 ‘a bachelor’ (Inglis 2007:8, shortened)

(19) *kuun*⁴ *saam* *koo*¹
 person three CLF
 ‘three people’ (Inglis 2007:11, shortened)

(20) Bengali

cho -*ṭa* *boi*
 six -CLF book
 ‘six books’ (David 2015:136)

- (21) *boi cho -ta*
 book six -CLF
 ‘the six books’ (David 2015:137)

While further parallels and matches in morphosyntactic pattern are needed in order to make a conclusive case for a shift from a language specifically like Khamti to Eastern Indo-Aryan, the striking differences from other Indo-Aryan languages, as well as the dynamics of change shown in the evolutionary scenario that we infer, lend support to the idea that the typological profile of Eastern Indo-Aryan languages is due to contact rather than diachronic trends realized elsewhere in Indo-Iranian.

Iranian languages on the other hand appear to have developed classifiers more frequently; classifiers emerge on nine independent branches within Iranian. Roughly half of these developments are preceded by stages with optional plural marking; ultimately, the state of affairs characterized by the GSS generalization appears to have been more active in Iranian than in Indo-Aryan, although if cognitive pressure is partly responsible for the development of classifiers in Iranian, it has most likely interacted with areal pressure, leading to the spread of this feature across languages.

The historical record makes it relatively clear that the development of full-fledged numeral classifiers in Persian was preceded by a prolonged period of optional plural marking, a state of affairs recapitulated by our SCM procedure. Our evolutionary model suggests that this was the case on branches leading to Wakhi; the Pamir languages Sariqoli, Shughni and Ishkashimi; Taleshi; and Kumzari as well. Hence, these developments are compatible with the claims made by the GSS hypothesis. However, contact may have also played a role; the proximity of Iranian languages to Turkic languages possessing classifiers cannot be ignored. As mentioned previously, Turkic and Iranian languages came increasingly into contact in post-Sasanian times. However, numeral classifiers are a likely to be a relatively late development, post-dating the onset of the Turkic expansion from about the 5th century CE (Yunusbayev et al. 2015). The Old Turkic (ca. 7th to 11th century) corpus contains no sortal numeral classifiers (Erdal 2004:226); however, the later literary language Chagatay (ca. 14th to 19th century) in Central Asia already exhibited certain classifiers, such as *baş* ‘head’ (Bodrogligeti 2001:155), and classifier use in modern Turkic languages is widespread, found among different branches of Turkic, including Kipchak (e.g., Tatar, Chen Zongzhen and Yi Liqian 1986:70), Oghuz (e.g., Turkmen, Clark 1998:169), and Karluk (e.g., Uzbek, Beckwith 1998). It is not clear whether this distribution reflects a widespread, relatively late diffusion of the feature under question throughout Turkic or the realization of some sort of drift or slant-like tendency on one hand, or the inheritance of a chronologically deep feature within the Turkic family. Complicating matters, Turkic classifiers are often Iranian loans; only in restricted occurrences do Iranian languages borrow Turkic classifiers (e.g., Sariqoli ← Uyghur). Further research and more methodological development is needed to address the question of whether Turkic classifiers are due to Iranian influence, Iranian classifiers are due to Turkic influence, or both groups developed classifiers as a response to loosening restrictions on plural marking under the influence of widespread multilingualism.

In some cases, the presence of numeral classifiers in Iranian languages is clearly due to

influence from other Iranian languages. The numeral classifiers found in Wakhi, Yaghnobi and certain Pamir languages are clearly identifiable as Tajik, given Tajik's importance as a lingua franca in the region where these languages are spoken.

In other cases, the role of contact is unclear. Kumzari, located in Oman and is separated from other Iranian languages by the Persian Gulf, exhibits the classifiers *-ta* and *-kas* (for human beings), formally identical to the Persian classifier *tā* and the Persian word *kas* 'person'. Kumzari is genetically very close to Old, Middle, and modern Persian (Skjærvø 1989), though it is not clear that it is a descendant of Old or Middle Persian. Given the relative isolation of Kumzari, it is possible that it developed these classifiers in parallel with Persian due to drift, but contact with another Iranian language is also a possibility, since there has been longstanding migration to Oman from the other side of the Persian gulf (Barth 1983).⁸ Our phylogenetic model suggests that Kumzari developed obligatory plural marking after a stage of optional plural marking; this may be due to contact with Arabic, which has obligatory plural marking.

Remaining Iranian languages with classifiers such as Mazandarani, Taleshi, and Gilaki share a core group of classifiers that are formally near-identical to Persian ones (e.g., *tā*), but as in the case of Kumzari, it is difficult to determine whether these forms are inherited or borrowed on the basis of sound change; at the same time, the dominance of Persian over these languages is well established (Borjian 2009), lending circumstantial evidence to the notion that they are Persian loans. Some Northwest Iranian languages have classifiers not found in Persian, e.g., Taleshi *gəla* (Paul 2011, Stilo 2018), perhaps cognate to Judeo-Tati *gile* 'time, instance' (Authier 2012:310). Many of these languages appear to have undergone a long-term trend toward optional plural marking. It is likely that classifiers spread through contact, but did so because of their utility and the cognitive enhancement they brought about.

From the evidence surveyed, a picture emerges where in the majority of cases, Indo-Iranian classifiers developed as both a response to optional plural marking AND contact, and this was realized both by means of pattern and matter borrowing. Iranian languages use borrowed matter as numeral classifiers, regardless of whether plural marking is optional or obligatory, as shown in Table 1. In some cases actual classifiers are borrowed (e.g., Zazaki *teney* from Turkish); in others, borrowed matter seems to have been repurposed as classifiers (e.g., Persian uses Arabic loanwords as classifiers in addition to inherited material, though they probably did not enter the language as such). However, Iranian languages with obligatory plural marking seem not to have developed classifiers from inherited matter. Yaghnobi has borrowed classifiers from Tajik. Pashto has likely borrowed the classifier *tana* (m.)/*teni* (f.) from a Turkic language, though it has been in the language long enough to be integrated into the gender system and participate in a phonological process. The case of

⁸Though the historical phonology of Kumzari is poorly understood, there is some evidence that it preserves the consonant of Old Iranian final **-aka-*, given the "etymologically latent *k*" in the definite form *mark-ō* (van der Wal Anonby 2015:38) < **marta-ka-*. Modern Persian does not preserve this consonant (cf. *tā* < Middle Persian *tāg* 'piece', most likely going back to an element **tāka-*). However, Kumzari seems to show preservation only in contexts where the form is suffixed, making it impossible to determine whether the loss of *-k* in *-ta* is regular or reflects a Persian borrowing.

	Borrowed matter	Inherited	Both
+clf,-obl.pl	Wakhi	Bhojpuri, Awadhi, Maithili, Nepali, Oriya, Bengali, Assamese, Bakhtiyari	Sariqoli, Ishkashimi, Zazaki, Modern Persian, Dari, Gilaki, Taleshi, Mazandarani,Ormuri
+clf,+obl.pl	Agia Varvara Romani, Pashto, Yaghnobi	Sinhala, Kumzari	

Table 1: Table showing whether numeral classifiers are based on borrowed or inherited matter in languages with numeral classifier systems.

Kumzari is ambiguous, as it is not clear whether its classifiers are Persian cognates or loans, or whether plural marking was obligatory in Kumzari when classifiers developed.

In Indo-Aryan, the majority of languages have developed classifiers from inherited matter, likely at a historical stage with optional plural marking. Romani, which has obligatory plural marking, is the only Indo-Aryan language to borrow classifiers. Sinhala is exceptional in having an inherited classifier (*denaa* < *jana*-) and obligatory plural marking.⁹

Ultimately, languages with non-obligatory plural marking have classifiers based on either inherited or borrowed matter, but languages with obligatory plural marking tend to have classifiers based solely on borrowed matter (except for Sinhala and maybe Kumzari). If languages of the latter type borrowed classifiers during a stage when plural marking was not optional, it was possibly because the prestige of the source language trumped the fact that numeral classifiers would bring no understandable cognitive benefit to a language with obligatory plural marking. Languages where the development of classifiers brought about a cognitive benefit as well may have borrowed their classifiers from source languages with and without relative prestige, and this is reflected by the combination of matter (pointing to prestige) and pattern (pointing to stable multilingualism) borrowing that they display. Further research at a wider phylogenetic scale is needed to conclusively address some of these questions.

⁹Modern Sinhala obligatory number marking is thought to have developed from an earlier stage where singular marking was non-obligatory (Nitz and Nordhoff 2010). While it is not possible from the available historical record to date the development of the *denaa* classifier relative to the development of a more restrictive number marking system, it is possible that the former development took place while singular marking still exhibited a degree of optionality. Either way, neither a system with obligatory number marking nor a system with optional singular marking should give rise to numeral classifiers under the Greenberg-Sanches-Slobin hypothesis.

7 Conclusions

Our findings suggest that the emergence of classifiers is tied to optional plural marking, partially explainable a statistical universal principle in line with the Greenberg-Sanches-Slobin generalization. At the same time, we find that classifiers emerged specifically under contact, reflecting local history and the contingencies of migration and trade. These results seem contradictory when one approaches the distribution of linguistic structures from the popular view that conceptualizes universal pressure and areal histories as conflicting, confounding, and competing factors. The contradiction is resolved, however, if we adopt the view from Distributional Typology (Bickel 2015) where emphasis is placed on the *interaction* between universal and areal factors (Bickel 2017).

From this perspective, the borrowing of classifiers (as matter or pattern) is not the mere product of historical contingency. Instead, this borrowing is driven by the decay in number marking, i.e. it follows the Greenberg-Sanches-Slobin generalization. Such a scenario explains our finding that classifiers were considerably more likely to be borrowed in the absence than in the presence of number marking. Areal effects alone cannot explain this difference, while universal effects alone cannot explain why classifiers entered through borrowing rather through spontaneous developments (e.g. by reanalyzing nominal juxtapositions, as in Hackstein's 2010 theory).

While our study supports a scenario of an interaction between universal and areal effects, we caution that our simplified coding scheme may not have picked up all relevant factors and that further research is needed to consolidate our conclusions. What is arguably the most urgent extension is a more fine-grained coding that captures the distribution of number marking over specific noun types, controlling for potential effects of a referential scale.

Appendix: state codings

Agia Varvara Romani [vlax1238] (+clf,obl.pl) Singular forms are always morphologically distinct from their plural counterparts, usually via the addition of a plural suffix or alternation of the final vowel (Iglá 1996:23ff.). When a numeral greater than one modifies a noun, the noun is morphologically plural (p. 45). When the counted item consists of indefinite objects, then *-tane* (< Turkish *tane* 'piece, part') appears next to the numeral; anaphoric use of this classifier is obligatory (p. 45).

Assamese [assa1263] (+clf,opt.pl) Assamese has a large inventory of sortal classifiers. Plural marking is optional (Borah 2012, Chowdhary 2012).

Avestan [aves1237] (-clf,obl.pl) Plural number is consistently marked, given rich agreement morphology (Hoffmann and Forssman 2004).

Awadhi [awad1243] (+clf,opt.pl) Classifiers are present; information regarding plural marking is difficult to extract from Saksena 1971:115ff., but it appears to be optional.

Bactrian [bact1239] (-clf,opt.pl) In late Bactrian, case and number distinctions have been neutralized due to the loss of distinctions between final vowels, resulting in “an unmarked form without ending ... which may be used with either sg. or pl. reference, and a marked pl. form” (Sims-Williams 2007:40).

Bagri [bagr1243] (-clf,morph.pl) There are at least three declensional classes, in one of which the plural and singular direct forms are identical. The distinction between animacy and inanimacy, rather than ending in *-i* versus other segments, is made by the author, who explicitly states that the suffix *-ā* is optional on animate nouns. This is almost akin to a mixture of a system like that of Hindi, where plural cannot be marked on some noun case forms, and a system where plural marking is truly optional.

Bakhtiyari [bakh1245] (+clf,opt.pl) Classifiers are present, and plural marking is variable (Anonby and Asadi 2014).

Bengali [beng1280] (+clf,opt.pl) Bengali has a large repertoire of numeral classifiers (David 2015:135). Classifiers are obligatory with non-numeric quantifiers and lower numbers; optional with numbers ending in ‘hundred’, ‘thousand’, ‘lakh’, etc. (p. 142). Numeral classifiers cannot cooccur with nouns denoting a countable unit, e.g., units of weight, currency, time, except in certain emphatic contexts (p. 142). Plural marking is non-obligatory (p. 76).

Bhojpuri [bhoj1244] (+clf,opt.pl) Numeral classifiers are present, and plural marking is non-obligatory (Tiwari 1960:120, 228, 230)

Dari [dari1249] (+clf,opt.pl) Classifier use is common, but not obligatory; plural marking is optional; it is not clear if plural marking can co-occur with classifier use as in Standard Modern Persian (Kiseleva 1985:74-5; Ioannesjan 1999:58-9).

Dhivehi [dhiv1236] (-clf,opt.pl) For nonhuman nouns, plural marking is optional when plurality is clear from context (Gnandesikan 2017:59).

Domari [doma1258] (-clf,opt.pl) Plural marking on enumerated nouns interacts significantly with whether the numbers are inherited Indo-Aryan forms or borrowed from Arabic (see Matras 2012:97, 188ff.). Plural number is optionally marked on nouns modified by 2–3 (inherited numbers), obligatory on nouns modified by 4–10 (Arabic numbers), and optionally marked on nouns from 11 upward (Arabic numbers).

Dumaki [doma1260] (-clf,obl.pl) For virtually all nouns, the singular form is distinguishable from the plural form (Lorimer 1939:24ff.); this is achieved via suffixation or a stem alternation.

Gilaki [gila1241] (+clf,opt.pl) Rastorgueva et al. (2012) list several classifiers. Classifier use is optional when enumerating nouns, but appears to be quite common and obligatory in anaphoric use. In all examples given of counted nouns, there is no overt plural marking on the head noun (regardless of whether a classifier is present). Plural can otherwise be marked by means of certain suffixes.

Gujarati [guja1252] (-clf,opt.pl) Cardona (1965:66–7) refers to the plural marker *-o* as “optional,” and it seems to largely be omitted when nouns are modified by a number greater than 1; so-called “variable” nouns display a special “dependent stem form” when they are semantically plural, regardless of the presence of the suffix *-o*.

Hindi [hind1269] (-clf,morph.pl) Hindi shows four declensional classes; the details of number marking are different for each one. Certain noun types are paradigmatically non-exhaustive; plural marking not morphologically possible on C-final masculine nouns in direct case (Oberlies 2005).

Iron Ossetic [osse1243] (-clf,obl.pl) Plural number is marked on nouns by means of the suffix *-t-* (Thordarson 2009:117). In most contexts (except for contexts of enumeration, see below), use of *-t-* appears to be obligatory. When a noun is enumerated by a numeral greater than one, the noun is marked by the suffix *-i* (Digor), identical to the genitive suffix. According to Thordarson (2009:132), this suffix continues the Old Iranian plural suffix **-ah*. Nouns enumerated by numbers greater than one are always marked in a way that renders them distinct from singular nouns.

Ishkashimi [ishk1246] (+clf,opt.pl) Ishkashimi contains at least three classifiers; nouns modified by a numeral greater than 1 can appear in singular or plural form (Paxalina 1959:50).

Judeo-Tati [jude1256] (-clf,opt.pl) Plural is marked on nouns with the suffix *-ho* (Authier 2012:79). Overt plural marking on enumerated nouns seems virtually non-existent.

Kalam Kohistani [indu1241] (-clf,obl.pl) Kalam Kohistani achieves plural marking on a number of nouns via a vowel fronting process, which also is found in oblique forms of nouns, and appears to mark plural consistently on nouns (Baart and Sagar 2004:21). The word *khur* ‘foot’ may show variability in plural marking, but it is not clear from the data given.

Kalasha [kala1372] (-clf,opt.pl) Plural marking is optional (Petersen 2015:35–6).

Kashmiri [kash1277] (-clf,morph.pl) According to Wali and Koul (1996:190ff.): “plurals are formed from singular stems by vowel change, palatalization and suffixation. A few nouns stay invariant. Masculine plurals are formed differently than the feminine plurals.” Mass nouns, most body parts, and borrowed English nouns use the same forms in both the singular and the plural. Masculine nouns do not change for plurality if they have certain phonotactic properties or are borrowed from Hindi/Urdu and English with a final consonant.

Khotanese Saka [khot1251] (-clf,obl.pl) Plural number is consistently marked, given rich agreement morphology (Emmerick 1989).

Khovar [khow1242] (-clf,opt.pl) Plural marking appears to be optional on the basis of examples provided in Endresen and Kristiansen 1981.

Khwarezmian [khwa1238] (-clf,obl.pl) Plural is consistently marked (Durkin-Meisterernst 2009).

Konkani [konk1267] (-clf,morph.pl) Certain noun categories have identical singular and plural endings in the direct case; otherwise, plural is consistently marked (Almeida 1989:126ff.)

Kumzari [kumz1235] (+clf,obl.pl) From Thomas’s (1930) description, plural marking appears to be obligatory. Kumzari numerals are nearly identical to their Modern Persian cognates; however, from seven upwards, the Kumzari numerals all end in *-tā*, which is analyzed as a suffix. For human beings, a suffix *-kay* attaches to the number one *yek(kay)*; for two onward, the suffix *-kas* is used. According to a newer description, the numeral classifier *-tā* or *-ta* in Kumzari can also occur on numerals below seven (van der Wal Anonby 2015:47)

Luri [luri1257] (-clf,opt.pl) According to MacKinnon (2003), plural marking is the same as in Modern Persian. No information regarding numeral classifiers is provided.

Maithili [mait1250] (+clf,opt.pl) Maithili has at least two classifiers (Burghart 1992:v. 1, 117). The suffix *-sab(h)* is an optional plural marker; when added to nouns that are inherently plural (e.g., vegetables), takes on the meaning “X and such things.” Some other suffixes exist for reference to persons, used in formal speech (Burghart 1992:v. 1, 50-1).

Marathi [mara1378] (-clf,morph.pl) Plural number must be marked on semantically plural nouns, except where morphologically impossible, e.g., masculine kinship terms, certain loanwords (Pandharipande 1997:366–7). Emeneau (1956:11) claims that Marathi has a classifier *jaṇ/jaṇī* (f.) that appears “when nouns denoting persons are numerated by numerals higher than four (and optionally for two to four).” Lambert (1943:243) says the following: “When the numerals refer to persons, special forms are used instead of *don*, *tin*, *car*; to other numerals the word *zə n* (m. *zə n*, cf.fem. *zə na(n)*; f. *zə ni*, cr.fem. *zə ni(n)*) is usually added. This word is often added also to the special forms of *don*, *tin*, *car*.” No examples are given. Katenina (1963:50) gives examples of the special forms *doghe*, *tighe*, *ćaughe*, as well as the forms *jaṇ* (m.) and *jaṇī* (f.) ‘people’ which show the latter form as a head noun, but never in close apposition with another (head) noun. The interaction between the numerals and *jaṇ(ī)* is striking; however, other grammars gloss these special forms simply as ‘both’, ‘the three’, and ‘the four’ respectively (Dhongde and Wali 2009:59). These forms appear in Old Marathi as substantivized numerals, e.g., *he tighe bhāu* ‘these three were brothers’ (Tulpule 1963, apud Southworth 1962:425).

Marwari [marw1260] (-clf,morph.pl) There are at least three declensional classes, in one of which the plural and singular direct forms are identical. Plural number is marked on plural nouns, where morphologically possible (Gusain 2004:20, 29).

Mazandarani [maza1291] (+clf,opt.pl) Classifiers are present, and plural marking is optional (Nawata 1984:9–10).

Mewati [mewa1249] (-clf,morph.pl) There are at least three declensional classes, in one of which the plural and singular direct forms are identical. Plural number is marked on plural nouns, where morphologically possible (Gusain 2003:20, 29).

Middle Persian [pahl1241] (-clf,opt.pl) Middle Persian can mark plural with the suffixes *-hā* and *-ān*, but plural is frequently unmarked on plural nouns (Skjærvø 2009:223).

Modern Persian [midd1350] (+clf,opt.pl) Modern Persian has several numeral classifiers, the most basic and widespread of which is *tā*, optionally used with numbers larger than one (Windfuhr and Perry 2009:478). Plural marking is optional, but the noun being modified can be marked for plural number if it has specific reference (Mahootian 1997:195). Classifiers are obligatory in anaphoric use.

Nepali [nepa1254] (+clf,opt.pl) Nepali contains several numeral classifiers (Acharya 1991:100); plural number is marked with *-haru*; according to Acharya this marking is optional (pp. 98–9). From Acharya’s examples, *-haru* can co-occur with numeral classifiers (p. 100). According to Bhim Lal Gautam (p.c.), *-haru* is obligatory with human nouns; however, non-human nouns cannot co-occur with overt plural marking and a classifier.

Old East Rajasthani [dhun1238] (-clf,morph.pl) There are at least three declensional classes, in one of which the plural and singular direct forms are identical. Plural number is marked on plural nouns, where morphologically possible (Metzger 2003).

Old Persian [oldp1254] (-clf,obl.pl) Plural number is consistently marked, given rich agreement morphology (Kent 1951).

Oriya [oriy1255] (+clf,opt.pl) Oriya has several classifiers; plural marking is optional (Neukom and Patnaik 2003).

Ormuri [ormu1247] (+clf,opt.pl) According to Kieffer (2003:133), classifiers are used as they are in Dari.

Pali [pali1273] (-clf,obl.pl) Pali generally maintains a clear morphological distinction between singular and plural; although the nominative singular and plural of ā-stems fell together due to regular sound change, a secondary plural suffix came into use in order to distinguish between the two numbers (Oberlies 2001:150–1)

Palula [phal1254] (-clf,obl.pl) Plural is consistently marked on count nouns with one of five suffixes, which are accompanied in some cases by stem alternations (Liljegren 2016:103–4).

Panjabi [panj1256] (-clf,morph.pl) There are at least three declensional classes, in one of which the plural and singular direct forms are identical. Plural number is marked on plural nouns, where morphologically possible (Bhatia 1993:214–5).

Parachi [para1299] (-clf,opt.pl) Little information about interactions between numeral modification and number marking can be found in Kieffer (2009). According to Morgenstierne (1929:50), plural marking is optional, and rare when a numeral modifies the noun. No information regarding classifiers is given.

Parthian [part1239] (-clf,opt.pl) Durkin-Meisterernst (2014:272) describes a scenario for all of Middle West Iranian whereby plural marking is optional on all enumerated nouns, but animate nouns are more often marked for plural number than inanimates.

Pashto [pash1269] (+clf,obl.pl) Plural number appears to be consistently marked on Pashto nouns (Penzl 1955:45ff.); most paradigms are relatively complex and contain stem alternations. Numeral classifiers are possible, and co-occur with nouns marked for plural number, possibly with gender agreement (p. 82).

Prakrit [maha1305] (-clf,obl.pl) Plural number is consistently marked, given rich agreement morphology (Woolner 1928).

Rakhshani Baluchi [west2368] (-clf,morph.pl) Nouns can be marked for indefiniteness and singularity via the suffix *-e*; otherwise, there is no morphological distinction between singular and plural (Barker 1969:3ff.).

Sangesari [sang1315] (-clf,morph.pl) Classifiers do not exist in Sangesari. According to Azami and Windfuhr (1972:70ff.), plural is consistently marked on oblique nouns with the suffix *-uon*, but rarely on direct nouns, except for a restricted set of items.

Sanskrit [sans1269] (-clf,obl.pl) Plural number is consistently marked, given rich agreement morphology (Macdonell 1910).

Sariqoli [sari1246] (+clf,opt.pl) Classifiers are present, and plural marking is optional (Paxalina 1971).

Saurashtran [saur1248] (-clf,morph.pl) Plural number is marked on Saurashtran nouns by means of two suffixes, *-nu* and *-lu* (< Telugu) (Ucida 1979:45–6). When a numeral greater than one modifies a noun, the noun is plural, but certain non-human nouns (e.g., days, years) do not take plural form.

Shina [shin1264] (-clf,morph.pl) Plural appears to be consistently marked on count nouns (Schmidt et al. 2008).

Shughni [shug1248] (-clf,opt.pl) Plural marking is optional, but classifiers do not appear to be present (Zarubin 1960).

Sindhi [sind1272] (-clf,morph.pl) There are at multiple declensional classes, in one of which the plural and singular direct forms are identical. Irregular plurals can be found for kinship terms. Arabic words often have distinctive plurals borrowed from the source language. Plural number is marked on plural nouns, where morphologically possible (Egorova 1966:27–8).

Sinhala [sinh1246] (+clf,obl.pl) When animate nouns are modified by a numeral, the form of the numeral used is different from that which is used when an inanimate noun is modified by a numeral (Chandralal 2010:60). Plural marking is obligatory, when applicable.

Siraiki [sera1259] (-clf,morph.pl) Certain noun categories have identical singular and plural endings in the direct case; otherwise, plural is consistently marked (Shackle 1976)

Sogdian [sogd1245] (-clf,opt.pl) Some Sogdian heavy stem nouns show a form that is identical to the singular in plural contexts (Yoshida 2009:313).

Sorani Kurdish [cent1972] (-clf,opt.pl) According to Blau (1980:45-6), the simple form of a noun can have either a singular or plural reading. In general, plural number is marked with the suffix *-an*.

South Tati [esht1238] (-clf,opt.pl) According to Yar-shater (1969:78), “nouns modified by a numeral higher than one, or by an expression denoting plurality, are generally expressed in the plural in Chali [a particular dialect].” Occasionally, however, the singular is used. In the other dialects, normally the singular is used for enumerated nouns. Plural marking otherwise seems to be the norm.

Taleshi [taly1247] (+clf,opt.pl) Taleshi has a number of numeral classifiers, use of which is non-obligatory (Paul 2011:181-2); additionally, “any noun following a numeral phrase is generally in the singular.” Elsewhere, plural number is marked with a suffix that varies from dialect to dialect.

Torwali [torw1241] (-clf,morph.pl) According to Grierson (1929:34), if a noun ends in a vowel, it can take a plural-marking suffix *-e*; otherwise, the singular and plural forms are identical.

Wakhi [wakh1245] (+clf,opt.pl) Native Wakhi numeral forms are in competition with Tajik numeral forms (Grjunberg and Steblin-Kamenskij 1988:89-90). The word for twenty (*bist*) is borrowed from Tajik, but numerals 20-30 can combine *bist* with either Wakhi or Tajik forms in the digits place. A number of classifiers are borrowed from Tajik. Unlike the situation in Yaghnobi, these classifiers can be used with both Wakhi and borrowed Tajik numerals. Plural is marked with the suffix *-iš(t)* (p. 19), but this marking appears to be optional.

Yaghnobi [yagn1238] (+clf,obl.pl) According to Xromov (1972:21–2), when numbers two and upward combine with nouns, the noun is found in the oblique singular form; if measure terms are used, the measure term is marked for oblique singular. The numerative *ta*, borrowed from Tajik, can be used, but only with Tajik numerals. Outside of the context of enumeration, *-t/d* is consistently used to mark plural number on nouns.

Zazaki [dim11238] (+clf,opt.pl) According to Paul (1998:19ff.), a morphologically singular noun can be used in a generic sense, but nouns denoting a plurality, definite or indefinite, take the plural ending. Plural marking is non-obligatory. An apparent sortal classifier *teney* co-occurs with nouns marked both for singular and plural.

References

- Acharya, J. (1991). *A Descriptive Grammar of Nepali and an Analyzed Corpus*. Washington D. C.: Georgetown University Press.
- Acquaviva, P. (2004). The morphosemantics of transnumeral nouns. In G. Booij, E. Guevara, A. Ralli, S. Sgroi, and S. Scalise (Eds.), *Morphology and linguistic typology. On-line proceedings of the Fourth Mediterranean Morphology Meeting (MMM4)* <http://mmm.lingue.unibo.it/mmm-proc/MMM4>.
- Aikhenvald, A. Y. (2000). *Classifiers. A typology of noun categorization devices*. Oxford Studies in Typology and Linguistic Theory. Oxford: Oxford University Press.
- Aikhenvald, A. Y. and R. Dixon (1998). Dependencies between grammatical systems. *Language* 74, 56–80.
- Almeida, M. (1989). *A description of Konkani*. Panaji: Thomas Stevens Konkani Kendra.
- Anonby, E. and A. Asadi (2014). *Bakhtiari studies: phonology, text, lexicon*. Uppsala: Acta Universitatis Upsaliensis.
- Authier, G. (2012). *Grammaire Juhuri, Ou Judeo-Tat, Langue Iranienne des Juifs du Caucase de L'est*. Beitrage Zur Iranistik. Ludwig Reichert Dr.
- Azami, C. A. and G. Windfuhr (1972). *A Dictionary of Sangesari with a Grammatical Outline*. Tehran: Franklin Book Programs.
- Baart, J. L. and M. Z. Sagar (2004). *Kalam Kohistani Texts*. Islamabad: National Institute of Pakistan Studies, Quaid-i-Azam University.
- Barker, M. A. (1969). *A Course in Baluchi*. Montreal: Institute of Islamic Studies, McGill University.
- Barth, F. (1983). *Sohar: culture and society in an Omani town*. Baltimore: Johns Hopkins University Press.
- Barz, R. and A. Diller (1985). Classifiers and standardisation: some South and South-East Asian comparisons. *Papers in South-East Asian Linguistics* 9, 155–84.
- Bauer, B. (2017). *Nominal Apposition in Indo-European*. Berlin, Boston: De Gruyter.
- Beckwith, C. I. (1998). Noun specification and classification in uzbek. *Anthropological Linguistics* 40(1), 124–140.
- Bhatia, T. K. (1993). *Punjabi: A cognitive-descriptive grammar*. London: Routledge.
- Bickel, B. (2003). Belhare. In G. Thurgood and R. J. LaPolla (Eds.), *The Sino-Tibetan languages*, pp. 546 – 70. London: Routledge.

- Bickel, B. (2015). Distributional typology: statistical inquiries into the dynamics of linguistic diversity. In B. Heine and H. Narrog (Eds.), *The Oxford Handbook of Linguistic Analysis*, 2nd edition, pp. 901 – 923. Oxford: Oxford University Press.
- Bickel, B. (2017). Areas and universals. In R. Hickey (Ed.), *The Cambridge Handbook of Areal Linguistics*, pp. 40–55. Cambridge: Cambridge University Press.
- Bisang, W. (2002). Classification and the evolution of grammatical structures: a universal perspective. *STUF - Language Typology and Universals* 55, 289–308.
- Bisang, W. (2017). Classification between grammar and culture: a cross-linguistic perspective. In T. Pommerening and W. Bisang (Eds.), *Classification from antiquity to modern times*. Berlin: De Gruyter.
- Blau, J. (1980). *Manuel de Kurde (dialecte Sorani)*. Paris: Librairie C. Klincksieck.
- Bodrogligeti, A. J. E. (2001). *A grammar of Chagatay*. Munich: Lincom Europa.
- Bollback, J. P. (2006). SIMMAP: Stochastic character mapping of discrete traits on phylogenies. *BMC Bioinformatics* 7, 88.
- Borah, G. (2012). Classifiers in Assamese: their grammar and meaning chains. In G. Hyslop, S. Morey, and M. W. Post (Eds.), *North East Indian linguistics*, Volume 4, pp. 293–314. New Delhi: Foundation Books.
- Borer, H. (2005). *Structuring Sense, Volume I: In Name Only*. Oxford: Oxford University Press.
- Borjian, H. (2009). Median succumbs to Persian after three millennia of coexistence: Language shift in the Central Iranian Plateau. *Journal of Persianate Studies* 2, 62–87.
- Burghart, R. (1992). *Introduction to Spoken Maithili in the Social Context*. Heidelberg: Südasien Institut, Abteilung für Ethnologie. 3 vols.
- Cardona, G. (1965). *A Gujarati reference grammar*. Philadelphia: University of Pennsylvania Press.
- Carpenter, B., A. Gelman, M. D. Hoffman, D. Lee, B. Goodrich, M. Betancourt, M. Brubaker, J. Guo, P. Li, and A. Riddell (2017). Stan: A probabilistic programming language. *Journal of statistical software* 76(1).
- Cassirer, E. (1923). *Philosophie der symbolischen Formen. Erster Teil: Die Sprache*. Berlin: Bruno Kassierer Verlag.
- Chandralal, D. (2010). *Sinhala*. Amsterdam, Philadelphia: John Benjamins.
- Chatterji, S. K. (1926). *The Origin and Development of the Bengali Language*. Calcutta: Calcutta University Press.

- Chen Zongzhen and Yi Liqian (1986). *Tata'er jianzhi*. Peking: Minzu chubanshe.
- Chowdhary, R. (2012). On classifiers in Asamiya. In G. Hyslop, S. Morey, and M. W. Post (Eds.), *North East Indian linguistics*, Volume 4, pp. 267–291. New Delhi: Foundation Books.
- Clark, L. (1998). *Turkmen reference grammar*. Wiesbaden: Harrassowitz.
- Corbett, G. G. (2000). *Number*. Cambridge: Cambridge University Press.
- Croft, W. (1994). Semantic universals in classifier systems. *Word* 45(2), 145–171.
- David, A. B. (2015). *Descriptive Grammar of Bangla*. Berlin, Boston: De Gruyter Mouton.
- Dhongde, R. V. and K. Wali (2009). *Marathi*. Amsterdam, Philadelphia: John Benjamins.
- Doetjes, J. (2012). Count/mass distinctions across languages. In C. Maienborn, V. von Heusinger, and P. Portner (Eds.), *Semantics: An International Handbook of Natural Language Meaning. Volume 3*, pp. 2559–80. Berlin: De Gruyter.
- Dunn, M., S. J. Greenhill, S. C. Levinson, and R. D. Gray (2011). Evolved structure of language shows lineage-specific trends in word-order universals. *Nature* 473(7345), 79–82.
- Durkin-Meisterernst, D. (2009). Khwarezmian. In G. Windfuhr (Ed.), *The Iranian languages*, Chapter 6, pp. 336–376. London: Routledge.
- Durkin-Meisterernst, D. (2014). *Grammatik des Westmitteliranischen (Parthisch und Mittelpersisch)*. Wien: Verlag der Österreichischen Akademie der Wissenschaften.
- Egorova, R. P. (1966). *Jazyk Sindxi*. Moscow: Akademia Nauk SSSR.
- Emeneau, M. B. (1956). India as a linguistic area. *Language* 32, 3–16.
- Emeneau, M. B. (1965 [1980]). India and linguistic areas. In A. S. Dil (Ed.), *Language and Linguistic Area: Essays by Murray B. Emeneau*, pp. 126–66. Stanford, CA: Stanford University Press.
- Emmerick, R. (1989). Khotanese and Tumshuqese. In R. Schmitt (Ed.), *Compendium Linguarum Iranicarum*, pp. 204–229. Wiesbaden: Ludwig Reichert Verlag.
- Endresen, R. T. and K. Kristiansen (1981). Khowar studies. *Acta Iranica* 21, 210–243.
- Erdal, M. (2004). *A Grammar of Old Turkic*. Leiden: Brill.
- Felsenstein, J. (2004). *Inferring phylogenies*. Sunderland, MA: Sinauer Associates.
- Geiger, W. (1942). *Beiträge zur singhalesischen Sprachgeschichte*. Munich: Verlag der Bayerischen Akademie der Wissenschaften.

- Gelman, A. and C. R. Shalizi (2013). Philosophy and the practice of Bayesian statistics. *British Journal of Mathematical and Statistical Psychology* 66, 8–38.
- Gil, D. (1987). Definiteness, noun-phrase configurationality, and the count-mass distinction. In E. Reulen and A. ter Meulen (Eds.), *The Representation of (In)definiteness*, pp. 254–269. Cambridge: MIT Press.
- Gil, D. (2013). Numeral classifiers. In M. S. Dryer and M. Haspelmath (Eds.), *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology.
- Gnandesikan, A. E. (2017). *Dhivehi: The Language of the Maldives*. Berlin, Boston: De Gruyter Mouton.
- Golden, P. (2006). Turks and Iranians: an historical sketch. In L. Johanson and C. Bulut (Eds.), *Turkic-Iranian Contact Areas: Historical and Linguistic Aspects*, pp. 17–38. Wiesbaden: Harrassowitz.
- Greenberg, J. (1972). Numeral classifiers and substantival number: Problems in the genesis type. *Working Papers on Language Universals* 9, 1–39.
- Grierson, G. A. (1929). *Torwali : an account of a Dardic language of the Swat Kohistan*. London: Royal Asiatic Society.
- Grinevald, C. (2000). A morphosyntactic typology of classifiers. In G. Senft (Ed.), *Systems of nominal classification*, pp. 50–92. New York: Cambridge University Press.
- Grjunberg, A. L. and I. M. Steblin-Kamenskij (1988). *La langue wakhi*. Paris: Editions de la Maison des Sciences de l'Homme.
- Gusain, L. (2003). *Mewati*. München: Lincom Europa.
- Gusain, L. (2004). *Marwari*. München: Lincom Europa.
- Hackstein, O. (2010). *Apposition and nominal classification in Indo-European and beyond*. Vienna: Verlag der Österreichischen Akademie der Wissenschaften.
- Haspelmath, M. (2013). Occurrence of nominal plurality. In M. S. Dryer and M. Haspelmath (Eds.), *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology.
- Her, O.-S. (2017). Deriving classifier word order typology, or greenberg's universal 20a and universal 20. *Linguistics* 55(2), 265–303.
- Heston, W. L. (1980). Some areal features: Indian or Irano-Indian. *International Journal of Dravidian Linguistics* 9(1), 141–157.

- Hoffmann, K. and B. Forssman (2004). *Avestische Laut- und Flexionslehre* (2nd ed.), Volume 84 of *Innsbrucker Beiträge zur Sprachwissenschaft*. Innsbruck: Institut für Sprachwissenschaft der Universität Innsbruck.
- Huelsenbeck, J. P., R. Nielsen, and J. P. Bollback (2003). Stochastic mapping of morphological characters. *Systematic Biology* 52(2), 131–158.
- Igla, B. (1996). *Das Romani von Ajia Varvara: deskriptive und historisch-vergleichende Darstellung eines Zigeunerdialekts*. Wiesbaden: Harrassowitz.
- Imai, M. and D. Gentner (1997). A crosslinguistic study of early word meaning: Universal ontology and linguistic influence. *Cognition* 62, 169–200.
- Inglis, D. (2007). *Nominal structure in Tai Khamti*, Volume 312 of *Research Paper*. Payap University.
- Ioannesjan, J. A. (1999). *Geratskij dialekt jazyka dari*. Moscow: Izdatel'skaja firma "vostočnaja literatura" RAN.
- Katenina, T. E. (1963). *Jazyk maratxi*. Moscow: Izdatel'stvo vostočnoj literatury.
- Kent, R. (1951). *Old Persian*, Volume 33 of *American Oriental Series*. New Haven: American Oriental Society.
- Kieffer, C. (2003). *Grammaire de l'ormuṛī de Baraki-Barak (Lōgar, Afghanistan)*. Wiesbaden: Dr. Ludwig Reichert Verlag.
- Kieffer, C. M. (2009). Parachi. In G. Windfuhr (Ed.), *The Iranian Languages*, pp. 693–720. London, New York: Routledge.
- Kiseleva, L. N. (1985). *Jazyk Dari Afganistana*. Moscow: Izdatel'stvo Nauka (glavnaja redakcija vostočnoj literatury).
- Krifka, M. (1995). Common nouns: A contrastive analysis of Chinese and English. In G. Carlson and F. J. Pelletier (Eds.), *The Generic Book*, pp. 398–411. Chicago: University of Chicago Press.
- Krishnamurti, B. and J. P. L. Gwynn (1985). *A grammar of Modern Telugu*. Oxford: Oxford University Press.
- Kruschke, J. K. (2011). Bayesian assessment of null values via parameter estimation and model comparison. *Perspectives on Psychological Science* 6, 299–312.
- Kuz'mina, E. (2007). *The origin of the Indo-Iranians*. Leiden: Brill.
- Lambert, H. M. (1943). *Marathi Language Course*. London: Oxford University Press.

- Lazard, G. (1963). *La langue des plus anciens monuments de la prose persane*. Paris: Klincksieck.
- Lehmann, C. (2000). On the German numeral classifier system. In C. Schaner-Wolles, J. R. Rennison, and F. Neubarth (Eds.), *Naturally! Linguistic studies in honour of Wolfgang Ulrich Dressler presented on the occasion of his 60th birthday*, pp. 249–253. Torino: Rosenberg & Sellier.
- Lehmann, C. (2010). On the function of numeral classifiers. In F. Floricic (Ed.), *Essais de typologie et de linguistique générale. Mélanges offerts à Denis Creissels*, pp. 435–445. Lyon: École Normale Supérieure.
- Lehmann, T. (1993). *A grammar of Modern Tamil* (2 ed.), Volume 1 of *PILC Publication*. Pondicherry Institute of Linguistics and Culture, India: Pondicherry, India: Pondicherry Institute of Linguistics and Culture. Includes index Includes bibliographical references (p. [379]-381).
- Liljegren, H. (2016). *A grammar of Palula*. Berlin: Language Science Press.
- Lorimer, D. L. R. (1939). *The Ḍumāki language : Outlines of the speech of the Ḍoma, or Bērīcho, of Hunza*. Nijmegen: Dekker and van de Vegt.
- Lucy, J. A. (1992). *Grammatical Categories and Cognition. A case study of the linguistic relativity hypothesis*. Cambridge: Cambridge University Press.
- Macdonell, A. A. (1910). *Vedic Grammar*. Strassburg: K.J. Trübner.
- Mache, A. (2012). *Numeral Classifiers in Persian*. München: LINCOM.
- MacKenzie, D. N. (1971). *A Concise Pahlavi Dictionary*. London: Oxford University Press.
- MacKinnon, C. (2003). The dialect of Xorramābād and comparative notes on other Lor dialects. *Studia Iranica* 31, 103–138.
- Maddison, W. P. and R. G. Fitzjohn (2014). The unsolved challenge to phylogenetic correlation tests for categorical characters. *Systematic Biology* 64, 127–136.
- Mahootian, S. (1997). *Persian*. London: Routledge.
- Masica, C. P. (1991). *The Indo-Aryan languages*. Cambridge: Cambridge University Press.
- Matisoff, J. (1978). *Variational Semantics in Tibeto-Burman: the ‘organic’ approach to linguistic comparison*, Volume 6 of *Occasional Papers of the Wolfenden Society on Tibeto-Burman Linguistics*. Philadelphia: Institute for the Study of Human Issues.
- Matras, Y. (2002). *Romani – A Linguistic Introduction*. Cambridge: Cambridge University Press.

- Matras, Y. (2012). *A Grammar of Domari*. Berlin: De Gruyter Mouton.
- Metzger, M. (2003). *Die Sprache der Vakīl-Briefe aus Rājasthān*. Heidelberg: Ergon.
- Moazami, M. (2014). *Wrestling with the Demons of the Pahlavi Widēwdād*. Leiden & Boston: Brill.
- Morgenstierne, G. (1929). *Parachi andOrmuri*, Volume 1 of *Indo-Iranian Frontier Languages*. Oslo: Instituttet for Sammenlignende Kulturforskning, H. Aschehoug & Co. (W. Nygaard).
- Mukherjee, T. (1963). *The Old Bengali language and text*. Calcutta: University of Calcutta.
- Nawata, T. (1984). *Mazandarani*, Volume 17 of *Asian and African Grammatical Manual*. Tokyo: Institute for the Study of Languages and Cultures of Asia and Africa, Tokyo University of Foreign Studies.
- Neukom, L. and M. Patnaik (2003). *A grammar of Oriya*. Number 17 in ASAS. Universität Zürich.
- Nielsen, R. (2002). Mapping mutations on phylogenies. *Systematic Biology* 51(5), 729–739.
- Nitz, E. and S. Nordhoff (2010). Subtractive plural morphology in Sinhala. In J. Wohlgemuth and M. Cysouw (Eds.), *Rara & Rarissima*, Berlin, pp. 247–266. De Gruyter.
- Oberlies, T. (2001). *Pali: A Grammar of the Language of the Theravada Tipitaka. With a Concordance to Pischel's Grammatik der Prakrit-Sprachen*. Berlin: Walter de Gruyter.
- Oberlies, T. (2005). *A historical grammar of Hindi*. Graz: Leykam.
- Oldenberg, H. (1909). *Ṛgveda: Textkritische und exegetische Noten, erstes bis sechstes Buch*. Berlin: Weidmannsche Buchhandlung. Rpr. 1970. Nendeln, Liechtenstein: Kraus.
- Oranskij, I. M. (1963 [1977]). *Les Langues Iraniennes*. Translated by Joyce Blau. Paris: Klincksieck.
- Padma, G. and C. Vijayalakshmi (2011). An analysis of continuous time markov chains using generator matrices. *International Journal of Computer Applications* 35(10), 20–24.
- Pagel, M. (1994). Detecting correlated evolution on phylogenies: A general method for the comparative analysis of discrete characters. *Proceedings of the Royal Society of London B* 255, 37–45.
- Pagel, M. and A. Meade (2006). Bayesian analysis of correlated evolution of discrete characters by Reversible-Jump Markov Chain Monte Carlo. *The American Naturalist* 167(6), 808–25.
- Pandharipande, R. (1997). *Marathi*. London: Routledge.

- Paul, D. (2011). *A comparative dialectal description of Iranian Taleshi*. Ph. D. thesis, University of Manchester.
- Paul, L. (1998). The position of Zazaki among West Iranian languages. In N. Sims-Williams (Ed.), *Proceedings of the Third European Conference of Iranian Studies held in Cambridge, 11th to 15th September 1995. Part I: Old and Middle Iranian Studies*, Wiesbaden, pp. 163–177. European Conference of Iranian Studies: Dr. Ludwig Reichert Verlag.
- Paxalina, T. N. (1959). *Iškašimskij Jazyk*. Moscow: Nauka.
- Paxalina, T. N. (1971). *Sarykol'sko-Russkij Slovar'*. Moscow: Nauka.
- Penzl, H. (1955). *A grammar of Pashto: a descriptive study of the dialect of Kandahar, Afghanistan*. Washington D. C.: American Council of Learned Societies.
- Petersen, J. H. (2015). Kalasha texts – with introductory grammar. *Acta Linguistica Hafnien-sia* 47, 1–275.
- Quine, W. V. O. (1960). *Word and Object*. Cambridge, MA: Technology Press of the Massachusetts Institute of Technology.
- Rastorgueva, V. S., A. A. Kerimova, A. K. Mamedzade, L. A. Pireiko, and J. I. Edelman (2012). *The Gilaki Language*. Uppsala: Uppsala Universitet.
- Revell, L. J. (2012). phytools: An R package for phylogenetic comparative biology (and other things). *Methods in Ecology and Evolution* 3, 217–223.
- Ross, M. (1996). Contact-induced change and the comparative method: Cases from Papua New Guinea. In M. Durie and M. Ross (Eds.), *The comparative method reviewed: Regularity and irregularity in language change*, pp. 180–217. New York: Oxford University Press.
- Ross, M. (2007). Calquing and metatypy. *Journal of Language Contact* 1, 116–143.
- Saksena, B. (1971). *Evolution of Awadhi (a branch of Hindi)*. Delhi: Motilal Banarsidass.
- Sanches, M. and L. Slobin (1973). Numeral classifiers and plural marking: An implicational universal. *Working Papers on Language Universals* 11, 1–23.
- Schiffmann, H. F. (1999). *A reference grammar of spoken Tamil*. Cambridge: Cambridge University Press.
- Schmidt, R. L., R. Kohistani, and M. M. Zarin (2008). *A grammar of the Shina language of Indus Kohistan*. Wiesbaden: Harrassowitz.
- Shackle, C. (1976). *The Siraiki language of Central Pakistan: a reference grammar*. London: School of Oriental and African studies, University of London.

- Silverstein, M. (1976). Hierarchy of features and ergativity. In R. M. W. Dixon (Ed.), *Grammatical relations in Australian languages*, pp. 112–171. Canberra: Australian Institute of Aboriginal Studies.
- Simpson, A., H. L. Soh, and H. Nomoto (2011). Bare classifiers and definiteness: A cross-linguistic investigation. *Studies in Language. International Journal sponsored by the Foundation “Foundations of Language”* 35(1), 168–193.
- Sims-Williams, N. (2007). *Bactrian documents from Northern Afghanistan, Vol. 2: Letters and Buddhist texts*. London: The Nour Foundation in association with Azimuth Editions.
- Skjærvø, P. O. (1989). Languages of Southeast Iran. In R. Schmitt (Ed.), *Compendium Linguarum Iranicarum*, pp. 363–369. Dr. Ludwig Reichert Verlag.
- Skjærvø, P. O. (2009). Old Iranian. In G. Windfuhr (Ed.), *The Iranian languages*, pp. 43–195. London: Routledge.
- Southworth, F. (1962). Review of *A Grammar of Old Marathi* by Alfred Master. *Language* 46, 507–513.
- Stilo, D. (2018). Numeral classifier systems in the praxes-iran linguistic area. In W. B. McGregor and S. Wichmann (Eds.), *The Diachrony of Classification Systems*, pp. 135–164. Amsterdam: Benjamins.
- Tang, M. and O.-S. Her (to appear). Insights on the Greenberg-Sanches-Slobin generalization: Quantitative typological data on classifiers and plural markers. *Folia Linguistica*.
- Thomas, B. (1930). *The Kumzari dialect of the Shihuh tribe, Arabia and a vocabulary*. London: The Royal Asiatic society.
- Thomason, S. G. and T. Kaufman (1988). *Language contact, creolization, and genetic linguistics*. Berkeley & Los Angeles: University of California Press.
- Thordarson, F. (2008 [2009]). *Ossetic Grammatical Studies*. Number 48 in Veröffentlichungen zur Iranistik. Vienna: Verlag der Österreichischen Akademie der Wissenschaften.
- Tirkey, M. K. . B. (2017). *The Kurux Language: Grammar, Texts, and Lexicon*, Volume 08 of *Brill’s Studies in South and Southwest Asian Languages*. Leiden: Brill.
- Tiwari, U. N. (1960). *The Origin and Development of Bhojpuri*. Kolkata: The Asiatic Society.
- Toulmin, M. (2009). *From linguistic to sociolinguistic reconstruction: the Kamta historical subgroup of Indo-Aryan*. Canberra: Pacific Linguistics, Research School of Pacific and Asian Studies, The Australian National University.
- Tulpule, S. (1963). *Prācīn marāṭhī korīv lekh (Old Marathi inscriptions)*. Poona: Poona University Press.

- Turner, R. L. (1962–1966). *A comparative dictionary of Indo-Aryan languages*. London: Oxford University Press.
- Ucida, N. (1979). *Oral literature of the Saurashtrans*. Calcutta: Simant.
- van der Wal Anonby, C. (2015). *A grammar of Kumzari: A mixed Perso-Arabian language of Oman*. Ph. D. thesis, Rijksuniversiteit te Leiden.
- Wali, K. and O. N. Koul (1996). *Kashmiri*. London, New York: Routledge.
- Wendtland, A. (2009). The position of the Pamir languages within East Iranian. *Orientalia Suecana* 58, 172–188.
- Windfuhr, G. and J. J. Perry (2009). Persian and Tajik. In G. Windfuhr (Ed.), *The Iranian Languages*, pp. 416–544. London: Routledge.
- Woolner, A. C. (1928). *Introduction to Prakrit*. Delhi: Motilal Banarsidass.
- Xromov, A. L. (1972). *Jagnobskij jazyk*. Moscow: Nauka.
- Yar-shater, E. (1969). *A Grammar of Southern Tati Dialects*. Number 1 in Median Dialect Studies. The Hague: Mouton.
- Yoshida, Y. (2009). Sogdian. In G. Windfuhr (Ed.), *The Iranian languages*, Chapter 5, pp. 279–335. New York; London: Routledge.
- Yunusbayev, B., M. Metspalu, E. Metspalu, A. Valeev, S. Litvinov, R. Valiev, V. Akhmetova, E. Balanovska, O. Balanovsky, S. Turdikulova, D. Dalimova, P. Nymadawa, A. Bahmanimehr, H. Sahakyan, K. Tambets, S. A. Fedorova, N. Barashkov, I. Khidiyatova, E. Mihailov, R. Khusainova, L. Damba, M. Derenko, B. Malyarchuk, L. P. Osipova, M. I. Voevoda, L. Yepiskoposyan, T. Kivisild, E. Khusnutdinova, and R. Villems (2015). The genetic legacy of the expansion of Turkic-speaking nomads across Eurasia. *PLoS Genet* 11(4), 1–24.
- Zarubin, I. I. (1960). *Šugnanskije teksty i slovar'*. Moscow: Izdatel'stvo Akademij Nauk SSSR.