

A universal about adnominal number agreement

Ranko Matasović

Department of Linguistics, University of Zagreb, Croatia < rmatasov@ffzg.hr >

Adnominal agreement can be defined as agreement in the domain of the noun phrase (NP), whereby the modifiers of the noun agree in categories like gender, number and case with the head noun. In many languages, including English, the modifiers agree with the head noun in number, cf. the English opposition between *this book* and *these books*. Likewise, many languages use the unmarked form of the noun (usually identical to the singular form) with numerals higher than one in numeral constructions (cf. Turkish *iki kitap* ‘two books’, where *kitap* ‘book’ is the unmarked, singular form), while other languages, such as English, use the number-marked form of the noun in such numeral constructions (cf. English *two books*, with *books* in the plural). This paper uses a genetically and areally balanced sample of 100 languages to show that languages like Turkish, in which nouns are unmarked for number in numeral constructions (with numerals above ‘one’), tend not to have adnominal number agreement. Although there are a few counter-examples to this universal, the correlation between the two features (having nouns unmarked for number in numeral constructions and lacking adnominal number agreement) is statistically highly relevant. We also show that this typological universal is not a consequence of different areal biases in the distribution of the features involved, i.e. that it is not a contact phenomenon. Lastly, we discuss the theoretical relevance of our findings and their implications for Jan Rijkhoff’s theory of ‘Seinsarten’ (Rijkhoff 2002, 2008).

KEYWORDS: number agreement, numerals, noun phrases, set nouns, typological universals.

1. Introduction

This paper will propose a statistical implicational universal relating two logically independent features that any language may have: number marking in numeral constructions with numbers higher than ‘one’, and number agreement in the domain of the noun phrase (NP). If our arguments are correct, it will add a further specimen to the stock of Greenbergian implicational universals (see, e.g., Greenberg 1978, Comrie 1989, Croft 2012), but we also hope to contribute to the general theory and typology of morphosyntactic features. We will show that number differs in some respects from the other features involved in agreement, which target the adnominal modifiers such as demonstra-

tives, adjectives and articles, especially gender and case. However, the focus of this paper will be on empirical rather than theoretical issues, and we also intend to present some typologically interesting new data about numeral constructions and agreement systems.

We will start by distinguishing two types of numeral constructions. In some languages, such as Turkish (Altaic), nouns construed with a numeral (higher than ‘one’) in a noun phrase do not get plural marking, but rather remain unmarked (1). In contrast, languages such as English pluralize nouns in such constructions (2):

(1) *iki kitap*
two book
‘two books’

(2) *two book-s*
two book-PL
‘two books’

Note that, in Turkish, one cannot mark the noun for plural number in this construction (**iki kitap-lar*, lit. ‘two book.PL’), just as, in English, the noun cannot remain in the unmarked singular (**two book*), although Turkish nouns are otherwise marked for number when they refer to a plurality of objects (e.g. *kitap-lar* ‘books’, *at-lar* ‘horses’, *adam-lar* ‘men’, etc.). In this paper, we will use a sample of one hundred languages and attempt to show that languages such as Turkish (let us call them **Type A** for convenience) do not have adnominal number agreement with a probability that cannot be attributed to chance. On the other hand, languages like English (henceforth **Type B** languages)¹ may or may not have adnominal number agreement. We will define agreement as a systematic covariance between a semantic or formal property of one element (in this case the noun) and a formal property of another (in this case the number marking on nominal modifiers, Corbett 2006).² Adnominal agreement is agreement in the domain of the NP (Matasović 2014, 2018), i.e. it is the agreement of nominal modifiers (typically adjectives, articles, demonstratives and numerals) with the head noun. The most common categories involved in adnominal agreement are gender, number and case (the so-called ‘phi-features’), but adnominal agreement in other categories, such as definiteness, is also attested (Matasović 2018: 37-8). Adnominal number agreement can be illustrated by the contrast between *this book* (sg.) and *these books* (pl.) in English, where the demonstrative agrees in number with the head noun.³ Note also that a language can belong to the Type A and have number agreement on verbs (but not on adnominal modifiers), which is the case, e.g., in

Turkish (Čaušević 1996: 107):

(3a) *bu kitap*
this book
'this book'

(3b) *bu kitap-lar*
this book-PL
'these books'

(4) *adam-lar at-lar sev-iyor-lar*
man.NOM-PL horse.ACC-PL love-PRES-3PL
'Men love horses'

Lastly, note that Type A languages must have number as a grammatical category in the NP (though not necessarily on verbs). In Turkish, plural can be regularly marked not only on nouns, but also on demonstratives and adjectives, when they are not used as adnominal modifiers, e.g. *bun-lar* 'these', so it is not the case that there is no number agreement in (3b) because the plural form of the demonstrative is formally identical to the singular, or that the category of number does not apply to demonstratives.

The rest of this paper will be dedicated to showing that the numeral construction such as the one found in Turkish (1) is a good predictor of the absence of adnominal numeral agreement, i.e. that in a language like Turkish a structure similar to (3b), but with number agreement (**bu-lar kitap-lar*), is typologically strongly disfavoured. This amounts to saying that there is an implicational statistical universal: if, in a language, nouns are unmarked for number in numeral constructions (with numerals above 'one'), then that language is unlikely to have adnominal number agreement. This universal applies in all languages in which number is a grammatical category, and for all agreement targets in the NP – typically adjectives, demonstratives and articles.⁴

2. The classification of languages into types

The approach adopted here starts with the classification of languages into only two types (A and B, as defined in Section 1). However, one could object that this classification of language types is not exhaustive. Besides languages in which nouns are (or are not) marked for number in constructions with numerals, we also have to take into account languages with classifiers, in which numerals cannot be construed directly with nouns in an NP. Rather, a classifier must be used to specify the class of

objects being counted. Such a language is Thai (Daic; Smyth 2002: 33):

(5a) *lûuk sǎam khon*
 child three CLASS
 ‘three children’

(5b) **lûuk sǎam*
 child three

On the other hand, classifier languages are very much like Type A languages, because in such languages nouns are usually not pluralized when they are construed with numerals and classifiers in the NP (Sanches & Slobin 1973, Rijkhoff 2002: 29, 50).⁵ For example, in Mandarin Chinese (Li 1999: 77) the plural marker *-men* cannot be used on nouns modified by numerals and classifiers, hence *san-ge xuesheng* (three-CLASS student) ‘three students’ is grammatical, while **san-ge xuexheng-men* (three-CLASS student-PL) is not. Moreover, classifier languages almost never have adnominal agreement in any category, including number⁶ (Matasović 2018). In any case, the main syntactic test showing whether a language belongs to Type A or to Type B is applicable regardless of whether the language has classifiers or not: if the form of the noun unmarked for number is used with numerals above ‘one’ (even if classifiers are also used), the language belongs to Type A; if the noun must be pluralized (or put into dual or paucal) in such a construction, then it belongs to Type B. Therefore, we shall treat classifier languages (henceforth the ‘Classifier’ type) as a special sub-type of Type A, unless they use the number-marked form of nouns with numerals above ‘one’ in the NP (regardless of the presence of classifiers).⁷

However, classifier languages are not the only problem our typology of numeral constructions faces. The classification of languages into discrete types is made difficult by several other factors, which must be addressed in turn.

Firstly, in some languages, there is apparently no number marking whatsoever. Although such languages are rare, they do seem to exist,⁸ but they are not in our sample. In all languages of our sample there is at least some formal way to indicate that a noun has plural reference, for at least a part (sometimes a very small part) of the nouns in the lexicon. However, in some languages (especially in languages with classifiers) it is difficult to establish whether a grammatical morpheme expresses number, collectivity, or is a quantifier. For example, ‘plural words’ (Dryer 1989) in isolating languages such as Thai are called ‘quantifiers’ in some grammars (Smyth 2002: 132), but in this paper they were nevertheless treated as plural markers expressing number in the NP.⁹

Secondly, in some languages there are no numerals, or this word class is limited to just a few items. De Araujo (2004: 95) analyzes Sabanes (Nambikwaran) as lacking any numerals, claiming that the lexeme *bala* is best interpreted as a dual marker, and that the meaning ‘three’ can be construed by combining *bala* and the adjective *kata* ‘single’. In Manggarai, as well as in many other Australian languages, there are numerals only for the numbers from one to three (Merlan 1982). In this paper, we treated those languages just like languages such as English, in which numerals are a major word class.¹⁰

And, thirdly and most importantly, languages can simultaneously have the two types of constructions mentioned above, with some nouns taking number marking in numeral constructions, and other nouns lacking it. This is found, e.g., in Tamil (Dravidian), Ngiti (Nilo-Saharan) and Makalero (Timor). In such languages, there is a ‘general’ form of some nouns, which is inherently unmarked for number. The general form of nouns that have it is used in numeral constructions, and it is usually identical with the singular form,¹¹ while other nouns lack the general form and are obligatorily marked for number in numeral constructions. In most such languages, the nouns referring to objects higher on the animacy hierarchy generally mark number in numeral constructions, while those lower on the animacy hierarchy do not. The animacy hierarchy (see, e.g., Corbett 2000: 54-56 and *passim*) can be represented in the following manner, as a scale:¹²

human > animate > inanimate > mass

The animacy hierarchy predicts that, in any language, if only some nouns referring to a plurality of objects are obligatorily marked for number in numeral constructions, then these nouns will belong to semantic classes nearer the left end of the scale than the other nouns in the language. For example, in Wappo (Wappo-Yukian, Thompson *et al.* 2006: 9), NPs with inanimate nouns often do not show plural marking in numeral and quantifier constructions, while NPs with animate nouns are consistently marked for plural in such constructions (6-7):

- (6) *kahon pasakis*
box two
‘two boxes’

- (7) *chica hopok’a kofo-mela*
bear three big-PL
‘three big bears’

On the other hand, the animacy hierarchy cannot account for all of the variation found in languages that have both number-marked and unmarked nouns in numeral constructions. Factors like definiteness play a role as well. A language may mark the number in definite NPs, while indefinite NPs are unmarked for number. This is the case, e.g., in Basque (8a-b) (Zubiri 2000: 65-7):

(8a) *hiru etxe-etan*
three house-INESS.PL
'in the three houses'

(8b) *hiru etxe-tan*
three house-INESS
'in three houses'

In our database, there are 16 languages in which number is marked on a subset of nouns (or the NPs they head) in numeral constructions. Such languages were classified as a special type, 'Split', different from both Type A and Type B. The 'Split' category includes 13 languages in which the split is based on semantic oppositions of animacy or humanness, and three languages (Basque, Brahui and Maori) in which it is based on the pragmatic feature of definiteness. It could be objected that this confuses the issue rather than clarifying it, since animacy and humanness are inherent features in the lexical representation of nouns, while definiteness depends on the context and the communicative intention of the speaker. However, it is clear that a language such as Basque cannot be classified as either a Type A or Type B language, because it would be arbitrary to consider its indefinite NPs as somehow more basic than its definite NPs (or vice versa). Hence, by classifying it as a Split language we are simply saying that it cannot fit into either of the otherwise well-defined types.¹³

The last question to be addressed in this section is whether there are languages in which number marking is completely optional. That is, are there languages in which number may, but need not be marked in NPs, and there is no semantic or pragmatic rule determining when number marking is obligatory? Greenberg (1978) has claimed that, in languages with optional number marking, the singular form of nouns may be used on nouns in constructions with numerals higher than 'one'. This universal is certainly relevant to our research, since it relates the optionality of number marking with the antecedent part of our universal implicature. However, the problem is that criteria for determining whether number marking is optional in a language are rather vague. In the languages included in our sample, we have found that animacy and

definiteness are generally very good predictors of number marking, if it is not obligatory in all NPs. It remains to be seen if number marking is ever truly optional, or if there are always factors determining the conditions of its use, but if languages with completely optional number marking exist, our universal probably does not say anything about them.

3. Problems with the classification

Although intuitively clear, the classification of languages into discrete types solely on the basis of the expression of number in numeral constructions may appear problematic for three reasons. Firstly, a language may have a special plural form of nouns used only when the noun is modified by a numeral. Although very rare, this type of construction is found in Djifanghor Nyun (North Atlantic branch of Niger-Kordofanian). In that language, indefinite pronouns and numerals agree with the noun in the plural by adding the suffix *-oŋ*; the plurality is also marked by the class prefix (Quint *forthcoming*: 12):

- (9) *kɛn-tédah-oŋ* *kan-nák-oŋ*
 GM-kitchen-PL GM-two-PL
 ‘two kitchens’

Nyun is not in our sample, but if it were, we would have classified it as a Type B language. This is because it uses the non-singular form of the noun in numeral constructions (with numerals higher than ‘one’), although its plural form used in the numeral construction differs from the one used in other constructions. Moreover, the plurality is indicated on both the noun and the numeral by gender markers (*kɛn-/kan-*) which differ for singular and plural referents.

Secondly, we must note that there are languages in which there is no absolute rule about the use of unmarked/singular forms of nouns modified by numerals. In Hausa, for example, “The noun is as a rule in the singular: thus, *shekara bakwai* = seven years where *shekara* ‘year’ is sg. The plural form may however also be occasionally used with an attributive numeral: *mutane biyu* ‘two men’, where *mutane* is pl.” (Smirnova 1982: 38). Apparently, the use of singular or plural in Hausa numeral constructions cannot be predicted from the animacy hierarchy. In a similar manner, Irish (Celtic branch of Indo-European) is inconsistent in its use of the (unmarked) singular and (marked) plural in the numeral construction: it uses the singular of most nouns (e.g. *trí bád* ‘three boats’, rather than **trí báid*, with the plural of *bád* ‘boat’), but the plural of the nouns denoting

units of measurement (*trí bliana* ‘three years’), body parts (*trí cosa* ‘three legs’, money units (*trí pingine* ‘three pence’), etc. In German, too, one can say *zwei Glas Wein* ‘two glasses of wine’ or *zehn Pfennig* ‘ten pennies’, where *Glas* and *Pfennig* are in the unmarked (singular) form (the respective plurals are *Gläser* and *Pfennige*).¹⁴ Otherwise, German uses the plural form of nouns in numeral constructions (e.g. *zehn Männer* ‘ten men’, with *Männer* being the plural of *Mann* ‘man’). Note that the universal that we propose does not imply that languages with such constructions cannot have adnominal number agreement – both Irish and German have it – since the large majority of numeral constructions in Irish and German use the marked (plural) form of the noun in numeral constructions (with numerals higher than ‘one’).

Furthermore, a language may have two different numeral constructions that behave differently with respect to the number of nouns (or the NPs that contain them). In Kryz (NE Caucasian), nouns are uninflected for number when they are construed with preposed numerals (higher than ‘one’), but they are in the plural when the numeral is postposed (10-11) (Authier 2009: 177-179):

(10) *šib ixlat*
 three story
 ‘three stories’

(11) *kil-arbi q'wa-d*
 arm-PL two-NEUT
 ‘two arms’.

In such languages, in which the simple criterion of juxtaposition of a numeral (higher than ‘one’) to a noun cannot clearly show whether the language belongs to the Type A or to the Type B, additional criteria for classification must be used, i.e. it must be checked which construction is basic, in the sense that it is more common and/or less pragmatically marked. In Kryz, for example, we may assume that the construction with the postposed numeral is basic, since in it the numeral carries a gender marker showing that it is a part of the same NP as the noun. If the numeral is preposed, there is no agreement in gender between the numeral and the noun, and the construction represented in (10) can better be understood as the equivalent of English ‘the stories, the three (of them)’ with the two elements in apposition. In Irish, one could argue that the use of the marked (plural) form is basic in numeral constructions, since the use of the unmarked (singular) form is limited to just a few semantically defined classes of nouns. Of course, problematic cases will remain, as grammars are often messy. If neither construction can be proved to be more basic, it is safest to classify the language as belonging to the Split type.

The classification of languages in the sample is represented in Table 1 below.¹⁵

TYPE	DEFINITION	EXAMPLES	NO. OF LANGUAGES IN THE SAMPLE
TYPE A (excluding Classifier languages)	Languages in which nouns (and NPs they head) are unmarked for number in numeral constructions	Kabardian Igbo Kayardild	36
Classifier languages (a sub-type of Type A)	Languages in which nouns quantified by a numeral have to be construed with a classifier ²³	Japanese Tzeltal Mokilese	17
TYPE B	Languages in which nouns (and NPs they head) are marked for number in numeral constructions (with numerals above 'one')	English Songhai Kryz	31
SPLIT LANGUAGES	Languages in which only a subset of nouns (and the NPs they head) is marked for number in numeral constructions	Tamil Manggarai Hup	16

Table 1. The typology of languages according to the patterns in numeral constructions

4. The sample

Our sample of 100 languages is a sub-set of the 300 languages sample we used in the monograph on the areal typology of agreement systems (Matasović 2018). It was constructed according to the same principles in order to have an areally and genetically unbiased selection. Languages were chosen so that the percentage of languages from different macro-areas in the sample roughly corresponds to that macro-area's overall linguistic diversity in terms of the number of languages. However, the number of languages from the Americas was slightly increased, because the genetic diversity of languages in the Americas (in terms of the number of distinct families) is significantly greater than in the other macro-areas. Similarly, Africa is slightly under-represented

in the sample (in terms of the number of languages), because only four indigenous language families are found on that continent.¹⁶

Table 2 shows the number of languages from particular macroareas in our sample, and the percentage of languages of the world in that macro-area (data were taken from < www.ethnologue.com >):

MACRO-AREA	NUMBER OF LANGUAGES IN THE SAMPLE	% OF THE WORLD'S LANGUAGES (ACCORDING TO ETHNOLOGUE)
Africa	20	30 %
Papunesia	16	19 %
Eurasia	40	36 %
Americas	24	15 %

Table 2. The areal distribution of languages in the sample

Care was also taken to include several languages from large families (Indo-European, Sino-Tibetan, Niger-Kordofanian, Afro-Asiatic, Austronesian), but that languages from those families in the sample should belong to different branches. Map 1 shows the languages in our sample.



Map 1. Languages in the Sample

Finally, here are the families from which languages were included in our sample: Afro-Asiatic (5 languages), Alor-Pantar, Altaic (6 languages), Araucanian, Athabascan (2 languages), Austro-Asiatic (2 languages), Austronesian (3 languages), Cariban (2 languages), Chapacura-Wari,

Dravidian (2 languages), Eskimo-Aleut, East Bougainville, Garrwan, Guaicuruan, Indo-European (10 languages), Iroquoian, Kartvelian, Khoesan, Lower Sepik, Maku, Mayan, Mirndi, Mosetenan, Moviman, Muskogean, Nambikwaran, Ndu, NE Caucasian (4 languages), Niger-Kordofanian (10 languages), Nilo-Saharan (4 languages), NW Caucasian (2 languages), Oto-Manguean, Pama-Nyungan (4 languages), Sino-Tibetan (6 languages), Solomons-East Papuan, Tai-Kadai, Tangkic, Timor, Totonacan, Tupian, Uralic (2 languages), Uto-Aztecan (2 languages), Wappo-Yukian, Witotoan, Yenisean, as well as four isolates. The full list of languages in the sample can be found in the Appendix.

5. The results

Our investigation has shown that languages belonging to the Type A are unlikely to have adnominal number agreement. In other words, we have found that the following implicational universal can be upheld with higher probability than can be attributed to chance: If a language has (only) nouns (and NPs they head) unmarked for number in numeral constructions, then it lacks number agreement in the NP.

In our database, there are 53 languages classified as languages of the Type A. The number of languages with at least some adnominal number agreement in our sample is 43 (in two languages – Cupeño and Kunuz Nubian – adnominal number agreement is marginal).¹⁷ Hence, the probability that a language should both belong to Type A and have adnominal number agreement is the probability that it has adnominal number agreement multiplied with the probability that it belongs to Type A, i.e. $0.53 \times 0.43 = 0.2279$. This means that 23 such languages are expected *a priori* in our sample of 100 languages, and only 8 are found: Bora, Chechen, Hungarian, Finnish, Georgian, Khwarshi, Kunuz Nubian and Wambaya.¹⁸ For this distribution, the chi-square value¹⁹ is 12.705, and the *p*-value is < 0.001 , which is highly statistically significant. Even if we exclude languages with classifiers from the Type A (for which, as we saw above, there is no compelling reason), there would still be only 8 languages having both adnominal number agreement and number unmarked forms of nouns in numeral constructions, but 16 would be expected *a priori*. For this distribution, the chi-square value would be 4.762, and the *p*-value 0.029, which would again be statistically significant if the threshold of significance is set to 0.05, as is customary in statistics.

If we lump together languages of the Split type and languages of the Type A, i.e. if we treat as a single type all languages in which at

least some nouns are unmarked for number in numeral constructions (as well as those languages in which all nouns are unmarked for number in such constructions), this type is still, at least statistically, incompatible with adnominal number agreement. There are 69 languages of this type (16 Split languages and 53 Type A languages) and 15 of them have adnominal number agreement (30 would be expected *a priori*). For this distribution, the chi-square value is 17.647 and the *p*-value is < 0.001, which is highly statistically significant.

Finally, let us look at the languages of the Split type in isolation. Of the 16 languages classified as the Split type, the majority (9) lack adnominal number agreement, while the remaining 7 have it (in one language, Wappo, adnominal number agreement is marginal). This distribution is not statistically significant (the *p*-value is 1, and the chi-square value is 0), as it is exactly what we would expect *a priori*. This means that, if a language has both numeral constructions with number marking of nouns, and those lacking it, one cannot predict anything with respect to the presence or absence of adnominal number agreement: a Split language may have it or not have it, it can be either way.

Table 3 summarizes our results and shows the statistical significance of the correlations between language types.

EXPRESSION OF PLURALITY	ADNOMINAL NUMBER AGREEMENT
All nouns (and the NPs they head) are unmarked for number in numeral constructions	NO (8 exceptions); $X^2 = 12.705$, $p < 0.001$
Some or all nouns are unmarked for number in numeral constructions (Type A and Split type lumped together)	NO (15 exceptions); $X^2 = 17.647$, $p < 0.001$
Split type	YES (no statistically relevant correlation)

Table 3. Summary of the results

At the end of this section, recall again that our operational definition of the Type A involved only the numeral construction, i.e. languages were classified as belonging to the Type A if they used a number-unmarked form of nouns when modified by numerals (higher than ‘one’) in the NP. We have seen above that eight Type A languages in our sample have some adnominal number agreement, and thus represent counter-examples to our universal implication. However, on closer inspection, most of those counter-examples are problematic, either

because adnominal number agreement in them is marginal, or because their attribution to Type A is not quite certain. Thus, Hungarian is classified as a Type A language, but in Hungarian, although the unmarked (singular) form of the noun is used in numeral constructions, only two lexical items, the demonstratives *az* and *ez*, show adnominal number agreement with the head noun in the NP (12). Adjectives and other modifiers (e.g. the definite article *a*) do not agree in number, so this is clearly a marginal agreement pattern (Matasović 2018).

- (12) *ez-ek a könyv-ek*
 this-PL ART book-PL
 ‘these books’

In a similar manner, Wambaya (an Australian language of the West Barkly family) has been classified as belonging to the Type A, since in that language the singular form of the noun also expresses ‘general number’ used to refer to dual and plural referents (Nordlinger 1998: 76), and dual and plural are used only when it is not clear from the context that referents are a pair or a plurality of objects. Hence, the unmarked form of the noun appears with the numerals. Moreover, all modifiers in Wambaya agree with the head noun in gender, number and case, “where morphologically possible” (Nordlinger 1998: 130).

However, on closer inspection it could be argued that even Wambaya is not a clear counter-example to our universal that Type A languages do not have adnominal agreement. Namely, all of the examples Nordlinger (1998) gives for nouns in the unmarked form having plural reference involve indefinite NPs, such as (13) (Nordlinger 1998: 73):

- (13) *Gaj-bi ng-a jigama*
 eat-NONFUT 1SG.A-NONPAST yam.GENDERIII(ACC)
 ‘I ate a/some bush yam(s)’

Moreover, if a demonstrative pronoun modifies the noun in a NP, thereby rendering it definite, then the noun is marked for number (Nordlinger 1998: 133):

- (14) *Ngajbi-g-a gujarrawulu marndag-bulu inuwuliyaga*
 see-3SG.A-PAST two(ACC) white_person-DU(ACC) that.GENDERI.DU(ACC)
 ‘He saw those two white men’

Note that in (14) the noun *marndag-bulu*, which is definite, carries the dual marker, although it is quantified by the numeral *gujarrawulu* ‘two’. Therefore, it may be that Wambaya actually belongs to the Split

type, and that the split is caused by a pragmatic feature of definiteness, i.e. that indefinite NPs in that language are unmarked for number in numeral constructions, while definite NPs are regularly marked for number in such constructions.²⁰ If that is indeed the case, Wambaya is not a true exception to our implicational universal. For reasons of space it is impossible to discuss in detail all the other six apparent counter-examples to our claim, but similar problems and uncertainties of classification can be observed in all of them.

6. *The areal distribution of language types*

In this section we will look at the areal distribution of the features we have studied in this paper. This is important, since statistical universals are sometimes accidental consequences of different areal biases in the global distribution of individual features related by universal implications. Thus, for example, it might well be that lack of adnominal number agreement and the use of nouns unmarked for number in numeral constructions (which defines our Type A) spread in the same areas due to prehistoric language contacts. If it were indeed the case, we would be in danger to mistake a contact phenomenon for a universal typological implication. Therefore, we need to check if the areal distributions of languages belonging to Type A and Type B in any way overlap with the distribution of languages with adnominal number agreement. Languages with classifiers, as is well known, show an areally biased distribution: they are concentrated in East Asia, but they are also found in South and Central America and in Oceania. In other parts of the world, they are very rare or non-existent (Map 2). Languages belonging to Type A, on the other hand, do not have a strongly areal distribution (Map 3): they seem to be fairly common in parts of West Africa, in North Australia (the Arnhem Land), as well as in parts of South America and Central Asia and the Caucasus. They are, on the whole, not particularly common elsewhere. The other types do not show a markedly areal bias in their distribution either: both Type B languages and the Split type languages are reasonably well attested everywhere.

In our other publications (Matasović 2014, 2018) we have shown that languages with adnominal agreement (including adnominal number agreement) have a biased areal distribution; Map 4 shows languages with adnominal agreement in our present sample. It will be seen that the areal distribution of such languages partially overlaps with the distribution of Type A languages (shown on Map 3): both types are common in



Map 2. Languages with classifiers in the sample (white pins)



Map 3. Type A languages (white pins). Classifier languages are excluded from Type A and represented in the same manner as the other languages on the map (black pins)

parts of West Africa, Northern Australia and Amazonia. Therefore, it is unlikely that the rareness of languages with both features is an accidental consequence of different areal biases of the two individual features they share.



Map 4. Languages with adnominal agreement (white pins)

7. Discussion

We have stated above (Section 1) that our distinction between Type A languages (such as Turkish) and Type B languages (like English) is reminiscent of Jan Rijkhoff's distinction between set noun languages and individual object languages (cf. Rijkhoff 2002, 2008). For Rijkhoff, set nouns do not denote individual objects, but rather sets, and those sets sometimes happen to have only one member. Hence in languages with set nouns such as Turkish, according to Rijkhoff, number is not a grammatical category as in languages with individual object nouns (such as English), in which every token of a noun inherently refers either to a singular object, or to a plurality of objects, and has to be marked for number. Therefore, individual object nouns are marked for number in numeral constructions as well, and in this way languages with individual object nouns are like our languages of the Type B. On the other hand, languages with set nouns are similar to our languages of the Type A, since in those languages what is often termed 'plural marking' is just an indication that the set of objects denoted by a noun has more than one member, and it does not occur in numeral constructions, since that would be redundant.

However, this similarity between Rijkhoff's types and ours is somewhat misleading, for two main reasons. Firstly, we have explicitly

included classifier languages as a sub-type of our Type A, since classifier languages generally do not mark nouns (or the NPs they head) for number in numeral constructions. In Rijkhoff's typology, this would not be acceptable, since classifier languages are in some respects fundamentally different from set noun languages. Rijkhoff's classification of noun types (or Seinsarten) is actually much more refined than our distinction between Type A and Type B languages. It relies on the presence or absence of the features 'shape' and 'homogeneity', which he holds for semantic universals (Rijkhoff 2002: 54). Thus, a general noun is defined by [-Shape] and designates a property that is characterized as not having a definite outline. Such nouns exist in certain languages, such as Yucatec (Mayan), with extensive use of classifiers, where the same lexical unit can mean very different things such as 'banana leaf' (*'un-wáal há'as*), 'banana fruit' (*'un-tz'út há'as*), 'bit of a banana' (*'um-p'út há'as*), and 'banana tree' (*'un-kúul há'as*), depending on the classifier used (Rijkhoff 2002: 47, with data from Lucy 1992: 74). A general noun is minimally characterized in terms of lexically coded information, and basically needs a classifier to specify its exact lexical content. On the other hand, a sort noun (defined by [-Shape, -Homogeneity]) designates a property characterized as not having a definite outline and not being agglomerative. Such nouns exist in languages with sortal classifiers, like Thai, where a classifier is not needed to specify the lexical content of a noun, so a noun with different classifiers can mean the same thing. We can see this by comparing, e.g., *mámûaŋ sām lûuk* (mango three CLASS.fruit) 'three mangoes' vs *mámûaŋ sām baj* (mango three CLASS.leaf) 'three mangoes', where both examples have the same meaning despite the use of different classifiers (Rijkhoff 2002: 49). The next item in Rijkhoff's classification are mass nouns, which probably exist in all languages. They are defined by the combination of features [-Shape, +Homogeneity] and define properties that do not have an outline and that are agglomerative, e.g. the noun for 'water' in most languages; set nouns have the feature [+Shape] and are underspecified for homogeneity, while individual object nouns denote objects that have a definite outline and are not agglomerative (e.g. the nouns for 'man', 'table' and 'stick' in familiar European languages). Thus, set nouns denote sets of objects, and the context shows when the set contains only one individual, while individual object nouns refer to individuals and must be marked for number (singular, dual, plural, or paucal, depending on the distinctions that a given language makes) when they refer to more than one individual. Lastly, collective nouns are defined as [+Shape, +Homogeneity] and designate properties that have a definite outline, but are agglomerative; they may be represented as separate groups of individuals (e.g. the nouns like children and family

in English) – such nouns exist in many languages. While Rijkhoff aims at a general typology of the ways nouns are conceptualized in languages, our typology of numeral constructions has no such ambition: it involves a single syntactic criterion to distinguish between two types of numeral constructions which, as we have seen, may co-exist in a single language.

Secondly, our Type A languages are different from Rijkhoff's set noun languages because we use a single defining criterion, while Rijkhoff defines his set noun type using three criteria which may, but need not coincide. These three criteria are:

1. having nouns unmarked for number construed with numerals above 'one' in the NP (the only criterion defining our 'Type A languages');
2. having all nouns inherently unspecified for number;
3. showing 'number discord',²¹ i.e. the use of the unmarked/singular form of verbs to agree with subjects denoting a plurality of referents, such as in the example (15) from Georgian (Kartvelian; Rijkhoff 2008: 742):

- (15) *sami kḡnuti goravs*
three kitten roll.3SG
'Three kittens are rolling'

In many languages these three criteria do coincide, but they do not have to. The examination of our sample has shown that all languages in which the noun is inherently unspecified for number use the number unmarked form of nouns with the numerals above 'one' in the NP, but the converse does not necessarily hold. In some languages number is, in principle, obligatorily marked on nouns that refer to a plurality of objects when used independently, but they are in the unmarked form when modified by a numeral above 'one' (e.g. in Hungarian). Likewise, many languages in which nouns are not marked for number in numeral constructions show 'number discord' (like Georgian), but there are also many which do not (e.g. Hungarian and Turkish).

To see why it is difficult to define a 'set noun language' in terms of necessary and sufficient conditions, it is instructive to take a closer look at Hungarian. At first sight, Hungarian looks like a typical set noun language: the equivalent of English 'two books' is *két könyv* (two book), and **két könyv-ek* (two book-PL) would be ungrammatical. However, in Hungarian, when nouns refer to a plurality of objects, they are normally marked for plural in most contexts, and if the verb is in the 3rd person plural form, its subject is always in the plural, as there is obligatory number/person agreement between the verb and its subject argument (Rounds 2001: 72):

- (16) *El-hallgat-t-ak* *a* *madar-ak* // **El-hallgat-t-ak a madar*
 VPREF-fall_silent-PRÉT-3PL ART bird-NOM.PL
 ‘The birds fell silent’

The example (16) shows that one cannot deduce from the context and the verb form (which is in the plural) that the reference of the number-unmarked form *madar* ‘bird’ is to a plurality of birds: rather, the noun referring to the subject must be in the plural, if it has plural reference. This is in sharp contrast with a more consistently set noun language such as Kabardian (NW Caucasian), in which the unmarked (formally singular) form of the noun in the subject position can co-occur with the plural form of a verb (17-18) (Matasović 2010; Kumakhov *et al.* 2006, I: 99):

- (17) *mal-əm* *yā-šx-ā-š*
 sheep-ERG 3PL-eat-PRÉT-AFF
 ‘The sheep have eaten’

- (18) *mal-xa-m* *yā-šx-ā-š*
 sheep-PL-ERG 3PL-eat-PRÉT-AFF
 ‘The sheep have eaten’

Therefore, Hungarian differs from Kabardian in the conditions for the use of nominal plural marking, but this does not mean that Hungarian is not a language of the set noun type. Besides the numeral construction, there are several other contexts in which Hungarian nouns used in the unmarked singular refer to a plurality of objects. For example, in existential constructions in which the existence or location of several referents is asserted, both the existential verb and its subject are normally in the singular (19):

- (19) *Van* *könyv* *a* *szobá-ban*
 there_is book(SG) ART room-INESS
 ‘There are books in the room’

Likewise, the singular is used after *hány* ‘how many’ (Rounds 2001: 90):

- (20) *Hány* *diák* *van* *a* *terem-ben?*
 how_many student(SG) there_is ART classroom-INESS
 ‘How many students are in the classroom?’

Hungarian can therefore be classified as a set noun language, although it is fair to admit that it is not such a typical token of this type as Kabardian, in which the unmarked/general form of the noun is more

consistently ambiguous as to the number of referents.

It is probably true that in most, if not in all languages, the singular form of a noun is used in general statements, which can be taken to refer to the set containing all objects that the noun can denote. In English, for example, the sentence *Man is mortal* is logically equivalent to *For all X, if X is a man, X is mortal*, or simply *All men are mortal*. In that sense, singular/unmarked nouns can have plural reference in nearly all languages.²² However, languages differ in terms of how often, or in which particular instances such general statements are made. For example, in English we would not use the singular of a noun in an existential construction as in Hungarian, i.e. (in standard British English) one cannot say **There is book in the room* in the sense ‘There are books in the room’. The fact that English is unlike Hungarian not just in its use of the plural form of nouns in the numeral construction, but also in other constructions that disallow singular nouns referring to a plurality of referents, makes it possible to firmly classify English as an individual object language, and Hungarian as a set noun language. Hungarian, on the other hand, is a set noun language just like Kabardian, but not such a prototypical token of this language type.

Therefore, we think that Rijkhoff’s typology of nominal aspect is best seen in terms of a scale, whereby languages with individual object nouns and those with set object nouns should not be defined by using sufficient and necessary conditions. Rather, both types of languages are characterized by clusters of features which tend to coincide, but not necessarily do. If a language uses the form of nouns unmarked for number in numeral constructions, it is also likely to use it in quantifier constructions and in existential constructions, but the converse does not hold. Likewise, languages with ‘number discord’ are likely to use nouns unmarked for number in numeral constructions, but again the converse does not hold.

We see, then, that Rijkhoff’s typology of nominal aspect should not be equated with our typology of numeral constructions. On the other hand, we believe that Rijkhoff’s theory of nominal aspect is relevant to typology of agreement systems, since it is based on the assumption that a certain type of nouns (the set nouns) is inherently unspecified for number. As such, set nouns cannot trigger number agreement in the NPs they head, since the modifiers of the noun cannot share with it the feature (number) that the noun itself lacks. Therefore, we expect languages with set nouns to lack adnominal number agreement. Since all prototypical set noun languages are very likely to be our Type A languages, the evidence we have collected, and presented in Section 5, appears to support this expectation.

To sum up: our empirical findings are formulated in a theory-neutral framework rather than in the framework of the theory of nominal aspect

proposed by Jan Rijkhoff. However, they can be interpreted so as to confirm that theory, which predicts that languages with set nouns will not have adnominal number agreement. Since Rijkhoff's prototypical languages with set nouns are our languages belonging to the Type A, this prediction is confirmed by our investigation.

8. Conclusion

In this paper we have shown that languages belonging to our Type A usually do not have adnominal number agreement. In other words, there is an implicational universal according to which adnominal number agreement does not tend to occur in languages in which nouns (and the NPs they head) are unmarked for number in numeral constructions (although they may be marked for number in other constructions). The statistical relevance of the correlation between belonging to Type A and not having adnominal number agreement is very strong indeed. Even if some languages have been accidentally misclassified, and the mistakes of classification do not cancel each other out, it is very likely that the correlation we have established would still hold.

As we mentioned in Section 1 above, adnominal number agreement is usually expressed by morphemes that also express gender and/or case. Because of this, it turns out that the type of numeral construction(s) a language has is also a good predictor of whether it has any adnominal agreement or not. That languages with classifiers are unlikely to have adnominal agreement has already been observed (Matasović 2018), and now we see that, more generally, languages in which nouns (or NPs they head) are unmarked for number in numeral constructions are also more likely than not to lack adnominal agreement. This is certainly an unexpected result that needs an explanation. Why do such languages lack adnominal number agreement? There is no logical necessity for this, as the exceptions in our sample illustrate.

Yet although possible, languages with adnominal number agreement and nouns unmarked for number in numeral constructions seem to be somewhat counter-intuitive: if number is not an inherent property of a noun that has to show up whenever it denotes a plurality of referents, how can it spread to other constituents of the NP? If we look at it that way, our implicational universal might be a simple corollary of the principle that languages cannot have agreement for features that are not inherent in the lexical representation of items that would trigger it. Unlike gender and case, number is not an inherent feature of the noun: hence, the feature of number cannot spread from the noun to the other elements of the NP. But, how-

ever logical it may sound, this claim certainly requires further reflection and research.

Abbreviations

A = agent; ACC = accusative; AFF = affirmative; ART = article; CLASS = classifier; DU = dual; ERG = ergative; F = feminine; FUT = future; GM = gender marker; INESS = inessive; M = masculine; NEUT = neuter; NOM = nominative; PL = plural; PRES = present; PRET = preterite; SG = singular; VPREF = verbal prefix.

Notes

¹ The uninspiring names Type A and Type B were chosen in order not to confuse the reader into assuming that the types are defined by a cluster of features rather than by a single criterion (having or lacking number marking in numeral constructions). As will be shown below, languages belonging to Type A are in some respects similar to languages with set nouns as defined by Jan Rijkhoff (2002), while those belonging to Type B are similar to his languages with individual object nouns. But Rijkhoff's typology of nominal aspect is much more complex than the one discussed here (see below, Section 7).

² Following Corbett (1991, 2000, 2006), we use the term *agreement* rather than *concord* (which is sometimes used as a term for agreement in the NP).

³ Unlike in English, in most languages with number agreement number morphemes are fused with morphemes expressing gender and/or case, e.g. in Croatian (gender + number + case), Welsh (gender + number) and Luiseño (case + number), see Matasović 2018. The system where modifiers agree with head nouns only in number is actually rather rare. It is found, e.g., in Kunuz Nubian and Gaahmg (both belonging to the Nilo-Saharan family; see Abdel Hafiz 1988 and Stirtz 2011, respectively), in Bobo (Mande, Dienst 2004) and a few other languages (see Matasović 2018 for discussion).

⁴ As we shall see below in Section 3, there are languages in which some, but not all nouns are unmarked for number in numeral constructions (with numbers higher than 'one'). In such languages the universal we propose does not apply. Note also that the universal is not limited to any particular set of adnominal modifiers as targets for agreement: it claims that if a language belongs to our Type A, then no adnominal modifiers will agree in number with the head noun.

⁵ There are exceptions, however, e.g. Baure (Arawakan; Krasnoukhova 2012: 121). In our sample, one such exception would appear to be Bora (Witotoan), which has gender/number agreement and a classifier system (Thiesen & Weber 2012). For other languages that have both classifiers and an agreement system (usually a gender system), or a mixed system with features of both classifiers and gender agreement, see Fedden & Corbett 2017.

⁶ In typological literature, there is an ongoing debate on whether the category of number is applicable to classifier languages. For example, Chierchia (2010) argues that classifier languages like Mandarin generally have mass nouns to which the category of number does not apply, while Nomoto (2013) offers a framework in which the category of number is equally applicable to languages with classifiers and those lacking them. The generally prevailing view now seems to be that, in most classifier languages (including Mandarin), not all nouns are mass nouns (Zhang 2013), hence

they can have the category of number.

⁷ In order to counter the objection that classifier languages were misclassified in this way, all the calculations of probabilities for typological correlations in Section 5 below were done separately for two cases: one, in which Type A languages included the classifier languages, and the other, in which classifier languages were excluded from Type A.

⁸ Pirahã (Mura; Everett 2005) is often cited as an example, although this has been disputed. Kwaza, an isolate spoken in Amazonia, does not express number, as a category, in the NP (Krasnoukhova 2012), but it is able to express the plurality of referents through its use of classifiers (Van der Voort 2004).

⁹ Note also that this implies that ‘number marking’ does not have to be morphological, i.e. that number does not have to be marked as an affix on the noun itself. The number marker can also be an independent word or an affix marking a modifier of the noun in the NP, e.g. the adjective or demonstrative. In this sense, number is better understood as marking the whole NP rather than its head, the noun itself.

¹⁰ In some languages, especially in the Amazon, underived numerals are adverbs or verb-like predicatives, and cannot occur as modifiers in the NP. However, apparently in all such languages numeral bases can be nominalized by affixation, and such nominalized numerals can then freely occur as parts of NPs (for examples see Krasnoukhova 2012: 115-117).

¹¹ In some languages, especially in the Cushitic branch of Afro-Asiatic, the general form of the noun is different from the singular, which has its own distinctive marking (Corbett 2000: 9-15).

¹² The Animacy hierarchy was originally formulated by Michael Silverstein (1976) in a different context (in a paper dealing with case marking in ergative languages). It has subsequently found applications in many aspects of grammar, including agreement systems. Our representation of the animacy hierarchy is somewhat simplified, in that it does not take into account personal pronouns, which play no role in adnominal agreement.

¹³ Note that such languages in no way contradict our statistical universal, since none of them have adnominal number agreement. The same holds for all of the languages of this type (i.e. with a pragmatically caused split in the treatment of nouns in numeral constructions) we came across in our research.

¹⁴ I owe the German examples to an anonymous reviewer of the first version of this paper. Even English uses the unmarked form of a few nouns with numerals, especially in non-standard idioms (e.g. *three quid* in certain varieties of British English), but on the whole they are exceptional.

¹⁵ As already mentioned, the typology does not include (and is not relevant to) languages in which there is no formal way to mark a noun that refers to a plurality of objects or a NP in which it occurs.

¹⁶ Those families are Niger-Kordofanian, Nilo-Saharan, Afro-Asiatic and Khoisan (which may actually represent an areal grouping of two or three families). Some sources, e.g. Glottolog <www.glottolog.org>, raise the number of families in Africa to over 50, but that is almost certainly overblown.

¹⁷ Marginal agreement patterns (Matasović 2018: 71-76) are those that are limited to just a few lexical items and/or those that occur in only a few infrequent constructions.

¹⁸ In Hungarian and Kunuz Nubian adnominal number agreement is marginal, limited to only a few lexical items. In Bora, Georgian and Chechen it is also rather limited: in Bora, only numerals show adnominal agreement; in Georgian, number/case agreement depends on the word order in the NP, and in Chechen only a handful of adjectives show gender/number agreement in the NP.

¹⁹ We used the chi-square calculator for goodness of fit, <www.socstatistics.com>.

²⁰ Rachel Nordlinger (personal communication) confirms that this interpretation is possible, although it cannot be proved without a detailed examination of Wambaya texts.

²¹ According to Rijkhoff (2002) this criterion is fulfilled in many, but not all set noun languages.

²² To take another example, in Ancient Greek, chiefly for stylistic reasons, singular form of nouns can be used to refer to a plurality of objects, e.g. *ho Mēdos* (ART.NOM.M.SG Mede.NOM.SG) ‘the Medes’, or *tō barbarikón* (ART.NOM.NEUT.SG barbarian.NOM.SG.NEUT) ‘the barbarians’ (Smyth 1984: 269). Otherwise, Ancient Greek is a typical individual object language, with nouns always in the plural (or dual) in numeral constructions, e.g. *pénte ánthrōpoi* (five man.NOM.PL) ‘five men’.

²³ For one language, Classical Nahuatl, the classification was problematic, as it shows features of both classifier languages and split languages (Sullivan 1983). In Classical Nahuatl, only animate nouns can have plural marking, but the language also obligatorily uses classifiers in numeral constructions. In this paper, it was tentatively classified as a classifier language.

Bibliographical References

- Abbott, Clifford 2006. *Oneida Teaching Grammar*. Green Bay: University of Wisconsin.
- Abdel-Hafiz, Ahmed S. 1988. *A Reference Grammar of Kunuz Nubian*. PhD dissertation. SUNY, Buffalo.
- Aikhenvald, Aleksandra 2008. *The Manambu Language of East Sepik, Papua New Guinea*. Oxford: Oxford University Press.
- Andrews, J. Richard 2003. *An Introduction to Classical Nahuatl*. Norman: University of Oklahoma Press.
- Andronov, Mikhail S. 1980. *The Brahui Language*. Moscow: Nauka.
- Authier, Gilles 2009. *Grammaire kryz, langue d’Azerbaïdjan*. Louvain: Peeters.
- Biggs, Bruce 1973. *Let’s learn Maori*. Auckland, NZ: Auckland University Press.
- Bradley, David 2003. Lisu. In Thurgood, George & LaPolla, Randy (eds.), *The Sino-Tibetan Languages*. London: Routledge. 222-235.
- Brauner, Sigmund 1995. *A Grammatical Sketch of Shona including Historical Notes*. Köln: Koppe.
- Broadwell, George A. 2006. *A Choctaw Reference Grammar*. Lincoln: The University of Nebraska Press.
- Carlson, Robert 1994. *A Grammar of Supyire*. Berlin: Mouton De Gruyter.
- Chierchia, Gennaro 2010. Mass nouns, vagueness and semantic variation. *Synthese* 174. 99-149.
- Comrie, Bernard 1989. *Language Universals and Linguistic Typology*. Chicago: The University of Chicago Press.
- Corbett, Greville 1991. *Gender*. Cambridge: Cambridge University Press.
- Corbett, Greville 2000. *Number*. Cambridge: Cambridge University Press.
- Corbett, Greville 2006. *Agreement*. Cambridge: Cambridge University Press.
- Courtz, Hendrik 2007. *A Carib Grammar and Dictionary*. Toronto: Magoria Books.
- Croft, William 2012. *Typology and Universals*. Cambridge: Cambridge University Press.

- Čaušević, Ekrem 1996. *Gramatika suvremenoga turskog jezika* [A Grammar of Contemporary Turkish]. Zagreb: Hrvatska sveučilišna naklada.
- De Araujo, Gabriel Antunes 2004. *A Grammar of Sabané, a Nambikwaran Language*. Utrecht: LOT.
- Derbyshire, Desmond 1979. *Hixkaryana*. Amsterdam: North Holland.
- Dienst, Stephan 2004. *Nominalmorphologie und Numerus im Bobo (Mande)*. Working Papers 34. Mainz: Johannes Gutenberg-Universität.
- Dixon, Robert M. W. 1972. *The Dyirbal Language of North Queensland*. Cambridge: Cambridge University Press.
- Dotton, Zura & Wagner, John D. 2016. *Chechen Grammar*. Manuscript.
- Dryer, Matthew S. 1989. Plural words. *Linguistics* 27. 865-895.
- Dumestre, Gérard 2003. *Grammaire fondamentale du bambara*. Paris: Karthala.
- Dum-Tragut, Jasmine 2009. *Armenian*. Amsterdam: Benjamins.
- Eaton, Helen 2010. *A Sandawe Grammar*. SIL International.
- Englund, Robert & Wolf, Werner 1960. *Finnische Sprachlehre*. Heidelberg: Groos.
- Epps, Patience 2008. *A Grammar of Hup*. Berlin: De Gruyter.
- Evans, Nicholas 1995. *A Grammar of Kayardild*. Berlin: Mouton de Gruyter.
- Everett, Daniel 2005. Cultural Constraints on Grammar and Cognition in Pirahã: Another Look at the Design Features of Human Language. *Current Anthropology* 46. 621-646.
- Everett, Daniel & Kern, Barbara 1997. *Wari'. The Pacaas Novos language of Western Brazil*. London: Routledge.
- Fedden, Sebastian & Corbett, Greville 2017. Gender and Classifiers in Concurrent Systems: Refining the Typology of Nominal Classification. *Glossa* 2,1. 1-47.
- Fiaga, Kwasi 1997. *Grammaire eve*. Paris / Lomé: Haho.
- Foley, William A. 1991. *The Yimas Language of New Guinea*. Stanford: Stanford University Press.
- Fortescue, Michael 1984. *West Greenlandic*. London: Croom Helm.
- Gaby, Alice R. 2006. *A Grammar of Kuuk Thaayorre*. PhD dissertation. University of Melbourne.
- Goossen, Irvy 2000. *Diné Bizaad: Speak, Read, Write Navajo*. Flagstaff: Salina Bookshelf.
- Gorelova, Liliya M. 2002. *Manchu Grammar*. Leiden: Brill.
- Gregores, Eva & Suárez, Jorge 1967. *A Description of Colloquial Guarani*. The Hague: Mouton.
- Greenberg, Joseph 1978. Generalizations about numeral systems. In Joseph Greenberg (ed.), *Universals of Human Language, Vol. III*. Stanford: Stanford University Press. 250-295.
- Grondona, Verónica María 1998. *A Grammar of Mocovi*. PhD dissertation. University of Pittsburgh.
- Harrison, Sheldon P. 1976. *Mokilese reference grammar. With the assistance of Salich Y. Albert*. Honolulu: University of Hawaii.
- Haspelmath, Martin 1993. *A Grammar of Lezgian*. Berlin: De Gruyter.
- Haude, Katharina 2006. *A Grammar of Movima*. Nijmegen: Nijmegen University.
- Heath, Jane 1999. *A Grammar of Koyra Chiini. Songhay of Timbuktu*. Berlin: Mouton de Gruyter.
- Heine, Bernd & Nurse, Derek (eds.) 2000. *African Languages. An Introduction*.

- Cambridge: Cambridge University Press.
- Hellwig, Birgit 2011. *A Grammar of Goemai*. Berlin: De Gruyter.
- Hewitt, George 1979. *Abkhaz*. London: Routledge.
- Hewitt, George 1996. *Georgian. A Learning Grammar*. London: Routledge.
- Hill, Jane 2005. *A Grammar of Cupeño*. Berkeley: University of California Press.
- Huber, Juliette 2011. *A Grammar of Makalero*. Utrecht: LOT.
- Jakovleva, Viktorija K. 1963. *Jazyk joruba*. Moscow: Nauka.
- Josephs, Lewis S. 1975. *Palauan Reference Grammar*. Honolulu: University of Hawaii Press.
- Kawachi, Kazuhiro 2007. *A Grammar of Sidaama (Sidamo), a Cushitic Language of Ethiopia*. PhD dissertation. SUNY, Buffalo.
- Khalilova, Zaira 2009. *A Grammar of Khwarshi*. Leiden: LOT.
- Kirsch, Beverley et al. 2004. *Teach Yourself Xhosa*. London: Hodder Arnold.
- Kite, Suzanne & Wurm, Stephen 2004. *The Duungidjawa language of Southeast Queensland: Grammar, Texts and Vocabulary*. Canberra: Australian National University.
- Klamer, Marian 2010. *A Grammar of Teiwa*. Berlin: Mouton de Gruyter.
- Klimov, G. A. & Edel'man, D. I. 1970. *Jazyk burušaski*. Moscow: Nauka.
- Koval', A. I. & Zubko, G. B. 1986. *Jazyk fula*. Moscow: Nauka.
- Krasnoukhova, Olga 2012. *The Noun Phrase in the Languages of South America*. Nijmegen: LOT.
- Kumakhov, M. A. et al. 2006. *Kabardino-čerkeskij jazyk*. Nalchik: Izdatel'skij centr 'Èl'-Fa'.
- Kung, Susan S. 2007. *A descriptive grammar of Huehuetla Tepehua*. PhD dissertation. University of Texas, Austin.
- Kuno, Susumu 1973. *The Structure of the Japanese Language*. Cambridge, Mass.: The MIT Press.
- Kutsch Lojenga, Constance 1993. *Ngiti. A Central-Sudanic Language of Zaire*. Köln: Rüdiger Köppe Verlag.
- Li, Yen-Hui Audrey 1999. Plurality in a classifier language. *Journal of East Asian Linguistics* 8. 75-99.
- Lucy, John A. 1992. *Grammatical categories and cognition: A case study of the linguistic relativity hypothesis*. Cambridge: Cambridge University Press.
- Macaulay, Monica 1996. *A Grammar of Chalcatongo Mixtec*. Berkeley: University of California Press.
- Matasović, Ranko 2010. *A Grammar of East Circassian (Kabardian)*. Manuscript available at <www.ffzg.hr/~rmatasov>.
- Matasović, Ranko 2014. Adnominal and verbal agreement: Areal distribution and typological correlations. *Linguistic Typology* 18,2. 171-214.
- Matasović, Ranko 2018. *An Areal Typology of Agreement Systems*. Cambridge: Cambridge University Press.
- Matisoff, James 1973. *A Grammar of Lahu*. Berkeley: University of California Press.
- Matthews, Stephen & Yip, Virginia 1994. *Cantonese: A Comprehensive Grammar*. New York: Routledge.
- McGregor, Robert S. 1986. *Outline of Hindi Grammar*. Oxford: Oxford University Press.
- Merlan, Francesca 1982. *Mangarayi*. London: Croom Helm.

- Meyer, Rudolf 1966. *Hebräische Grammatik* I-IV. Berlin: De Gruyter.
- Molochieva, Zarina 2010. *Tense, Aspect, and Mood in Chechen*. PhD dissertation. Leipzig University.
- Mushin, Ilana 2012. *A Grammar of (Western) Garrwa*. Berlin: Mouton de Gruyter.
- Nguyen, Dinh Hoa 1974. *Colloquial Vietnamese*. London: Routledge.
- Nomoto, Hiroki 2013. *Number in Classifier Languages*. PhD dissertation. University of Minnesota, Minneapolis / Saint Paul.
- Nordlinger, Rachel 1998. *A Grammar of Wambaya, Northern Territory (Australia)*. Canberra: Australian National University.
- Norman, Jerry 1988. *Chinese*. Cambridge: Cambridge University Press.
- N'un, Maun Maun 1963. *Birmanskij jazyk*. Moscow: Izdatel'stvo vostočnoj literatury.
- Onishi, Masayuki 2012. *A Grammar of Motuna*. München: Lincom Europa.
- Polian, Gilles 2006. *Éléments de grammaire du tsetal*. Paris: L'Harmattan.
- Quint, Nicolas *forthcoming*. Djifanghor Nyun (Casamance, Senegal). In Lüpke, Friederike (ed.), *Oxford Guide to the World's Languages: Atlantic*. Oxford: Oxford University Press.
- Rice, Keren 1989. *A Grammar of Slave*. Berlin: Mouton De Gruyter.
- Rijkhoff, Jan 2002. *The Noun Phrase*. Oxford: Oxford University Press.
- Rijkhoff, Jan 2008. On flexible and rigid nouns. *Studies in Language* 32,3. 727-752.
- Rounds, Carol 2001. *Hungarian. An Essential Grammar*. London: Routledge.
- Sakel, Jeanette 2002. *A Grammar of Mosetén*. Berlin: Mouton De Gruyter.
- Sanches, Mary & Slobin, Linda 1973. Numeral classifiers and plural marking: An implicational universal. *Working Papers on Language Universals* 11. 1-22.
- Schiffman, Harold F. 1999. *A Reference Grammar of Spoken Tamil*. Cambridge: Cambridge University Press.
- Seyoum, Mulugeta 2008. *A Grammar of Dime*. Utrecht: LOT.
- Silverstein, Michael 1976. Hierarchy of features and ergativity. In Dixon, Robert M. W. (ed.) *Grammatical Categories in Australian Languages*. Canberra: Australian Institute of Aboriginal Studies. 112-71.
- Smeets, Ineke 2008. *A Grammar of Mapuche*. Berlin: Mouton De Gruyter.
- Smirnova, Mirra A. 1982. *The Hausa Language*. Moscow: Nauka.
- Smyth, David 1995. *Colloquial Cambodian*. London: Routledge.
- Smyth, David 2002. *Thai: An Essential Grammar*. London: Routledge.
- Smyth, Herbert Weir 1984. *Greek Grammar*. Cambridge, MA: Harvard University Press.
- Sohn, Ho-Min 1999. *The Korean Language*. Cambridge: Cambridge University Press.
- Stirtz, Timothy 2011. *A Grammar of Gaahmg, a Nilo-Saharan Language of Sudan*. Utrecht: LOT.
- Sullivan, Thelma D. 1983. *Compendio de la gramática Náhuatl*. Ciudad de México: UNAM.
- Tamura, Suzuko 2000. *The Ainu Language*. Tokyo: Sanseido.
- Tereščenko, N. M. 1966. Èneckij jazyk. In Lytkin, V. I. & Majtinskaja, K. E. (eds.), *Jazyki narodov SSSR. Volume 3: Finno-Ugorskije jazyki i samodijskie jazyki*. Moscow: Nauka. 438-457.
- Terrill, Angela 1998. *Biri*. Munich: Lincom Europa.

- Thiesen, Wesley & Weber, David 2012. *A Grammar of Bora with Special Attention to Tone*. Dallas, TX: SIL International.
- Thompson, Sandra *et al.* 2006. *A Reference Grammar of Wappo*. Berkeley, CA: University of California Press.
- Thorne, David A. 1993. *A Comprehensive Welsh Grammar*. Oxford: Blackwell.
- Tucker Childs, Gordon 1995. *A Grammar of Kisi*. Berlin: De Gruyter.
- Vajda, Edward J. 2004. *Ket*. München: Lincom.
- Van Breugel, Seino 2014. *A Grammar of Atong*. Leiden: Brill.
- Van der Voort, Hein 2004. *A Grammar of Kwaza*. Berlin: Mouton De Gruyter.
- Vietze, Hans-Peter 1974. *Lehrbuch der mongolischen Sprache*. Leipzig: VEB.
- Wegener, Claudia 2012. *A Grammar of Savosavo*. Berlin: Mouton De Gruyter.
- Zhang, Niina Ning 2013. *Classifier Structures in Mandarin Chinese*. Berlin: Mouton De Gruyter.
- Zubiri, Ilari 2000. *Gramática didáctica del euskera*. Bilbo: Didaktiker, S.A.
- Zymeri, Isa 1993. *Colloquial Albanian*. London: Routledge.

APPENDIX

List of sampled languages (AA-N = adnominal number agreement; M indicates a marginal agreement pattern)

LANGUAGE	CODE	FAMILY	AA-N	NumberType	REFERENCES
Abkhaz	abk	NW Caucasian	NO	B	Hewitt 1979
Ainu	ain	Isolate	NO	A	Tamura 2000
Albanian	aln	Indo-European	YES	B	Zymeri 1993
Armenian (Eastern)	hye	Indo-European	NO	A	Dum-Tragut 2009
A'tong	aot	Sino-Tibetan	NO	Classifier	Van Breugel 2014
Bambara	bam	Niger-Kordofanian	NO	A	Dumestre 2003
Basque	eus	Isolate	NO	Split	Zubiri 2000
Biri	bii	Pama-Nyungan	NO	A	Terrill 1998
Bobo	bbo	Niger-Kordofanian	YES	B	Dienst 2004
Bora	bor	Witotoan	M	A/Classifier	Thiesen & Weber 2012
Brahui	brh	Dravidian	YES	Split	Andronov 1980
Burmese	mya	Sino-Tibetan	NO	Classifier	N'un 1963
Burushaski	bsk	Isolate	YES	B	Klimov & Edel'man 1970
Cantonese	cnt	Sino-Tibetan	NO	Classifier	Matthews & Yip 1994

A universal about adnominal number agreement

LANGUAGE	CODE	FAMILY	AA-N	NumberType	REFERENCES
Carib	car	Cariban	YES	Split	Courtz 2007
Chechen	che	NE Caucasian	YES	A	Molochieva 2010; Dotton & Wagner 2016
Choctaw	cho	Muskogean	NO	A	Broadwell 2006
Croatian	hrv	Indo-European	YES	B	
Cupeno	cup	Uto-Aztecan	M	B	Hill 2005
Dime	dim	Afro-Asiatic	YES	B	Seyoum 2008
Duungidjawan	duu	Pama-Nyungan	NO	A	Kite & Wurm 2004
Dyirbal	dbl	Pama-Nyungan	NO	Split	Dixon 1972
Enets	ene	Altaic	NO	A	Tereščenko 1966
English	eng	Indo-European	YES	B	
Ewe	ewe	Niger-Kordofanian	NO	A	Fiaga 1997
Finnish	fin	Uralic	YES	A	Englund & Wolf 1960
French	fra	Indo-European	YES	B	
Fula	fub	Niger-Kordofanian	YES	B	Koval' & Zubko 1986
Gaahmg	gaa	Nilo-Saharan	YES	B	Stirtz 2011
Garrwa (Western)	grr	Garrwan	NO	Split	Mushin 2012
Georgian	kat	Kartvelian	YES	A?	Hewitt 1996
Goemai	goe	Afro-Asiatic	NO	Split	Hellwig 2011
Greek	ell	Indo-European	YES	B	
Guarani	gua	Tupian	NO	A	Gregores 1967
Hausa	hau	Afro-Asiatic	YES	B	Smirnova 1982
Hebrew	heb	Afro-Asiatic	YES	B	Meyer 1966
Hindi	hin	Indo-European	YES	B	McGregor 1986
Hixkaryana	hix	Cariban	NO	A	Derbyshire 1979
Hungarian	hun	Uralic	YES	A	Rounds 2001
Hup	hpd	Maku	NO	Split	Epps 2008
Igbo	ibo	Niger-Kordofanian	NO	A	Heine & Nurse 2000

LANGUAGE	CODE	FAMILY	AA-N	NumberType	REFERENCES
Japanese	jap	Altaic	NO	Classifier	Kuno 1973
Kabardian	kbd	NW Caucasian	NO	A	Matasović 2010
Kayardild	gyd	Tangkic	NO	A	Evans 1995
Ket	ket	Yenisean	YES	B	Vajda 2004
Khalkha	khk	Altaic	NO	A	Vietze 1974
Khmer	khm	Austro-Asiatic	NO	Classifier	Smyth 1995
Khwarshi	khv	NE Caucasian	YES	A	Khalilova 2009
Kisi	kis	Niger-Kordofanian	YES	B	Tucker Childs 1995
Korean	kor	Altaic	NO	Classifier	Sohn 1999
Koyra Chiini	kch	Nilo-Saharan	NO	A	Heath 1999
Kryz	kry	NE Caucasian	YES	B	Authier 2009
Kuuk Thaayorre	thd	Pama-Nyungan	NO	A	Gaby 2006
Kwaza	kwz	Isolate	NO	Classifier	Van der Voort 2004
Lahu	lah	Sino-Tibetan	NO	Classifier	Matisoff 1973
Lezgian	lez	NE Caucasian	NO	A	Haspelmath 1993
Lisu	lis	Sino-Tibetan	NO	Classifier	Bradley 2003
Lithuanian	lit	Indo-European	YES	B	
Makalero	mjb	Timor	NO	Split	Huber 2011
Manambu	mle	Ndu	YES	B	Aikhenvald 2008
Manchu	mnc	Altaic	NO	A	Gorelova 2002
Mandarin	cmn	Sino-Tibetan	NO	Classifier	Norman 1988
Manggarai (Mangarayi)	myi	Mangarayi	YES	Split	Merlan 1982
Maori	mri	Austronesian	NO	Split	Biggs 1973
Mapuche (Mapudungun)	arn	Araucanian	NO	A	Smeets 2008
Mixtec (Chalcatongo)	mig	Oto-Manguan	NO	A	Macaulay 1996
Mocovi	mev	Guaicuruan	NO	B	Grondona 1998
Mokilese	mok	Austronesian	NO	Classifier	Harrison 1976
Moseten	mos	Mosetenan	YES	B	Sakel 2002
Motuna	mot	East Bougainville	YES	Split	Onishi 2012

A universal about adnominal number agreement

LANGUAGE	CODE	FAMILY	AA-N	NumberType	REFERENCES
Movima	mzp	Movima	NO	A	Haude 2006
Nahuatl (Classical)	nhn	Uto-Aztecan	NO	Classifier	Andrews 2003
Navajo	nav	Athabaskan	NO	A	Goossen 2000
Ngiti	nti	Nilo-Saharan	NO	Split	Kutsch Lojenga 1993
Nubian (Kunuz)	kzh	Nilo-Saharan	YES-M	A	Abdel-Hafiz 1988
Oneida	one	Iroquoian	NO	A	Abbott 2006
Palauan	pau	Austronesian	NO	Classifier	Josephs 1975
Sabanes (Sabané, Nambikuára)	nab	Nambikuaran	NO	A	De Araujo 2004
Sandawe	sad	Khoesan	YES	Split	Eaton 2010
Savosavo	svs	Solomons East Papuan	YES	B	Wegener 2012
Shona	shn	Niger-Kordofanian	YES	B	Brauner 1995
Sidamo	sid	Afro-Asiatic	YES	Split	Kawachi 2007
Slave	sla	Athabaskan	NO	A	Rice 1989
Spanish	spa	Indo-European	YES	B	
Supyire	sup	Niger-Kordofanian	YES	B	Carlson 1994
Tamil	tam	Dravidian	NO	Split	Schiffman 1999
Teiwa (Tewa)	twe	Alor-Pantar	NO	A	Klamer 2010
Tepehua (Huehuetla)	tee	Totonacan	NO	Classifier	Kung 2007
Thai	tha	Tai-Kadai	NO	Classifier	Smyth 2002
Turkish	tur	Altaic	NO	A	Čaušević 1996
Tzeltal (Oxchuc)	tzl	Mayan	NO	Classifier	Polian 2006
Vietnamese	vie	Austro-Asiatic	NO	Classifier	Nguyen 1974
Wambaya	wam	Mirndi	YES	A	Nordlinger 1998
Wappo	wao	Wappo-Yukian	M	Split	Thompson <i>et al.</i> 2006
Wari' (Chapacura)	pav	Chapacura- Wanham	NO	A	Everett & Kern 1997
Welsh	wel	Indo-European	YES	B	Thorne 1993

LANGUAGE	CODE	FAMILY	AA-N	NumberType	REFERENCES
West Greenlandic	kal	Eskimo-Aleut	YES	B	Fortescue 1984
Xhosa	xho	Niger-Kordofanian	YES	B	Kirsch <i>et al.</i> 2004
Yimas	yee	Lower Sepik	YES	B	Foley 1991
Yoruba	yor	Niger-Kordofanian	NO	A	Jakovleva 1963