

Languages with set nouns do not have adnominal number agreement



RANKO MATASOVIĆ
UNIVERSITY OF ZAGREB

Introduction



- Rijkhoff (2002): in languages with set nouns the unmarked noun is inherently neither singular nor plural, and the plural marking only resolves the ambiguity. In languages with individual object nouns the unmarked noun is inherently singular, and cannot be used to refer to more than one object.
- Oromo (Kushitic branch of Afro-Asiatic) *farda* ‘horse’ and ‘horses’, *nad’eeni* is both ‘woman’ and ‘women’, while in English *book* refers only to a single object, and *book-s* to a plurality of objects. Oromo nouns apparently denote sets rather than individual objects, and the context shows when the set contains just one individual.

Introduction



- Moreover, in languages with set nouns, such as Oromo, nouns construed with a numeral do not get number marking, but rather remain unmarked (1) ([Rijkhoff \(2002: 46\)](#)). In contrast, languages with individual object nouns, such as Dutch, mark the number on nouns in such constructions (2).



- (1) *gaala lamaani*



two camel



‘two camels’



- (2) *twee boek-en*



two book-PL



‘twoo books’

Introduction



- Rijkhoff's typology has consequences for our understanding of number agreement: this paper will show that languages with set nouns (almost) never have adnominal number agreement (on a sample of 100 languages).
- Agreement is defined 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).
- 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, cf. English *this book* (sg.) vs. *these books* (pl.).

Introduction



- The classification of languages into Set noun languages and Individual Object languages is not exhaustive: in languages with classifiers 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 (Smyth 2002: 33):
 - (3a) *lûuk sǎam khon*
 - child three CLASS
 - ‘three children’
 -
 - (3b) **lûuk sǎam*
 - child three
- Classifier languages are very much like set noun languages in that nouns are not pluralized when they are construed with numerals and classifiers in the NP (there are a few exceptions). Moreover, classifier languages almost never have adnominal agreement in any category, including number (but see Fedden and Corbett 2017).

The classification of languages into types



- The approach adopted here classifies languages into the following types: 1. languages with individual object nouns (the IO type); 2. languages with set nouns (the Set type); 3. languages with classifiers (the Class type). However, the classification of languages into discrete types is made difficult by several factors:
- Firstly, in some languages, there is no number marking whatsoever (e.g. Pirahã, Everett 2005). No such languages are in our sample.
- Secondly, in some languages there are no numerals, or this word class is limited to just a few items, e.g. Sabanes (Nambikwaran, De Araujo 2004: 95).
- Thirdly, in some languages, according to Rijkhoff (2008), nouns do not exist as a separate word class (e.g. in Samoan, an Austronesian language). No such languages are in our sample.

The classification



- And, lastly and most importantly, languages can simultaneously have both IO nouns and Set nouns, e.g. Tamil (Dravidian), Ngiti (Nilo-Saharan) and Makalero (Trans-New Guinea).
- In such languages, there is a general form of some nouns, usually identical with the singular form, while other nouns are obligatorily marked for plural (or dual, paucal) if they refer to a plurality of objects, and so they represent IO nouns. The distribution of IO and Set nouns can be partly predicted by the Anymacy hierarchy (Corbett 2000: 54-56 *et passim*):
 -
 - human > animate > inanimate > mass

The classification



- Wappo (Wappo-Yukian, Thompson et al. 2006: 19):
-
- (4) *Mansa:naʔ-i pina*
- apple-NOM few
- “The apples are few”
-
-
- (5) *Ceʔeʔ onoʔšiʔ-te*
- copula Indians-PL
- “They are Indians”
-
-
- Factors like definiteness and specificity play a role as well. A language may treat its nouns as individual object nouns only if they are definite, while indefinite nouns are treated as set nouns.

The classification



- In Basque, only definite nouns carry plural marking; indefinites can refer to individual objects as well to several objects (Zubiri 2000: 65-7):
- (6a) *Irakasle-a etorri da*
 - teacher-SG.DEF come AUX
 - “The teacher came”
 -
- (6b) *Irakasle bat etorri da*
 - teacher one come AUX
 - “A teacher came”
 -
- (6c) *Irakasle-ak etorri dira*
 - teacher-PL.DEF come AUX
 - “The teachers came”
 -
- (6d) *Zenbait irakasle etorri da*
 - some teacher come AUX
 - “Some teachers came”
 -



- (7a) *hiru etxe-etan*
- 3 house-INESS.PL
- ‘in the three houses’
-
- (7b) *hiru etxe-tan*
- 3 house-INESS
- ‘in three houses’
-
-
- In our database, there are 15 languages in which a formal distinction between singular and plural (sometimes also dual and paucal) is lacking in a part of the lexicon, while it exists (and is obligatory) in another part of the lexicon. Such languages were classified as a special type, “Split”, different from both Set noun languages (Set) and Individual object noun languages (IO).

Summary of the classification



Type	Definition	Examples	No. of languages in the sample
Individual object noun languages	Languages in which nouns cannot be used in the singular to refer to a plurality of objects	English, Songhai, Mosesten	42
Set noun languages	Languages in which nouns may be used in the singular (or general) form to refer to a plurality of objects	Kabardian, Igbo, Kayardild	26
Classifier languages	Languages in which nouns quantified by a numeral have to be construed with a classifier	Japanese, Tzeltal, Mokilese	17
Split languages	Languages in which only a subset of nouns may be used in the singular or general form to refer to a plurality of objects	Tamil, Manggarai, Hup	15

The sample



- Our sample of 100 languages is a sub-set of the 300 languages sample used in our previous work (Matasović 2018). The selection is meant to be areally and genetically representative.

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 %

Languages in the sample



The Results



- 1. If a language has (only) nouns with underspecified number, then it lacks number agreement in the NP. More informally, languages with (only) set nouns and those with classifiers do not have adnominal number agreement.
- In our database, there are 26 languages classified as languages with set nouns. Of these, only one (Wambaya, West Barkly) has adnominal number agreement. In that language, the singular form of the noun also expresses ‘general number’ used to refer to dual and plural referents (Nordlinger 1998: 108), 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. However, all modifiers agree with the head noun in gender, number and case, “whenever morphologically possible” (Nordlinger 1998: 180).

The Results



- In (8), the head noun is in the dual (Nordlinger 1998: 182):
- (8) *Ngarri-yulu gujarrawulu alag-ulu.*
- 1SG.POSS-DU(NOM) two(NOM) child-DU(NOM)
- ‘my two children’
-
- On closer inspection it could be argued that even Wambaya is not a clear counter-example to the universal claim that languages with set nouns do not have adnominal agreement. Namely, all of the examples Nordlinger (1998) gives for nouns in the unmarked form having plural reference involve indefinite and/or non-specific NPs, such as (9) (Nordlinger 1998: 109):
-
- (9) *Gaj-bi ng-a jigama*
- eat-NON.FUT 1SG.A-NON.PAST yam.GENDERIII(ACC)
- “I ate a/some bush yam(s)”



- Nordlinger (1998: 110) adds that “If it is necessary to clearly specify that a nominal is singular, the free form numeral *garndawuga*- ‘one’ can be used. The use of this numeral ensures that the general reading is not possible” (10):
-
- (10) *Garndawuga* *ngiy-a* *wankurradi* *marrgulu*
one.GENDERIV(ACC) 3SG.NON.M.A-PAST lay(NON.F) egg.GENDERIV(ACC)
- “She laid one egg”
- Therefore, it may be that Wambaya actually belongs to the Split type, and that the split is caused by a pragmatic feature of specificity, i.e. that non-specific nouns in that language are set nouns, while specific nouns are individual object nouns. If that is indeed the case, the universal we proposed is without exception in our sample.

The Results



- The number of languages with at least some adnominal agreement in our sample is 43. That means that the probability that a language should have both set nouns and number agreement is the probability that it has adnominal number agreement multiplied with the probability that it belongs to the Set type, i.e. $0.43 \times 0.26 = 0.1118$. This means that 11 such languages are expected *a priori*, and only one (Wambaya) is attested. For this distribution, the chi-square value is 10.215. The *p*-value is 0.001, which is statistically significant.
- There are also 17 languages classified as classifier languages, and only one (Bora) has limited number agreement on numerals in the NP. This distribution is again statistically significant: The chi-square value (for goodness of fit) is 5.53, and the *p*-value is 0.019, which is again statistically significant.



- Of the 15 languages classified as the Split type, the majority (9) lack adnominal number agreement, while the remaining 6 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*.
- Finally, if we lump together languages of the Split type and languages of the Set type, i.e. if we treat as a single type all languages in which at least some nouns are set object nouns, this type is still, at least statistically, incompatible with adnominal number agreement. There are 41 languages of this type (15 Split languages and 26 Set languages) and seven of them have adnominal number agreement (18 would be expected *a priori*). For this distribution, the chi-square value is 9.576 and the p -value is 0.002, which is statistically significant.

The Results



- 2. If a language has (only) nouns that may, but need not be specified for number, then a form unmarked for number is obligatorily used with numerals.
- This implication holds for all 43 languages classified as Set and Class in our database.
- The converse of the implicational universal (2) does not hold: a language can use the singular (or non-plural) form of nouns with its numerals, but belong to the IO type as defined above (i.e. its singular form of nouns cannot have plural reference in other contexts and constructions). In our sample, there are 10 languages of the IO type which use the singular form of nouns with the numerals: In Hausa and Mocovi, the use of non-plural forms of nouns with numerals is not without exceptions, and they were also classified as individual object languages.

The Results



- Note also that there are languages in which there is no absolute rule about the use of non-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).
- Finally, a language may have two different numeral constructions that behave differently with respect to the number of nouns. This is the case in Welsh, which has two constructions: with the preposition *o*, when the noun is in the plural, and without it, when it is singular. So, in Welsh one can say *naw o ddynion* ‘nine men’ (nine of men.pl), but also *dau afal* ‘two apples’ (two apple), Thorne 1993: 204.

The Results



- 3. If a language uses the singular form of nouns with its numerals in the NP, then it is unlikely to have adnominal number agreement.
- This means that, if languages with set nouns were defined by the criterion of having singular (or non-plural) form of nouns construed with numerals in the NP, the generalization that languages with set nouns do not have adnominal number agreement would still hold, although the correlation would not be as strong as if the other criterion is used. In our sample, there are 59 languages in which the singular form of the noun is used with numerals in the NP; there are 43 languages with adnominal number agreement, but only 8 languages that have both features (adnominal number agreement and the singular/non-plural form of nouns used with numerals). For this distribution, the chi-square value is 15.413, and the *p*-value is < 0.001 , which is statistically significant.

Summary of the results



Expression of plurality	Adnominal number agreement
All nouns in the sg. may be used to refer to a plurality of objects (the Set type)	NO (exception: Wambaya); $X^2 = 10.215$, $p = 0.001$
Some nouns in the sg. can be used to refer to a plurality of objects (the Set and Split type lumped together)	NO (7 exceptions); $X^2 = 9.576$, $p = 0.002$
Nouns in the sg. are construed with numerals above 1	NO (8 exceptions); $X^2 = 15.413$, $p < 0.001$
Classifiers	NO (possible exception: Bora); $X^2 = 5.53$; $p = 0.0019$
Split type	YES (no statistically relevant correlation)

The areal distribution of types



- Languages with classifiers in the sample (white dots):



The areal distribution



- Languages with set nouns (white dots):



The areal distribution



- In our previous work (Matasović 2014, 2018) we have shown that languages with adnominal agreement have a biased areal distribution (white dots represent languages with adnominal number agreement):



Conclusion



- There is an implicational universal according to which adnominal number agreement does not tend to occur in languages with nouns underspecified for number: if a language has nouns that can refer to a plurality of objects in their singular (or non-plural) form, then that language is unlikely to have adnominal number agreement.
- Languages with set nouns regularly use the singular form of nouns with their numerals, but the converse of this claim does not hold: some languages that do not belong to the “set noun” type use the singular (or non-plural) form of nouns in numeral constructions. However, languages that use the singular (or non-plural) form of nouns in numeral constructions are also unlikely to have adnominal number agreement.



- 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 languages with set nouns are also more likely than not to lack adnominal agreement.
- Why do languages with set nouns lack adnominal number agreement? There is no logical necessity for this, as we can clearly imagine a grammatical rule stating that, although the noun need not be marked for number, if it is so marked, then some or all of its modifiers in the NP have to be marked for number as well. But, with the probable exception of Wambaya in our sample, and maybe a few more languages, such rules seem not to be found in human languages.



- An imagined example of such a language: if *da* means ‘house’, *-k* is the plural suffix, and *pi* is the demonstrative stem, then both *dak pik* ‘these houses’ and *da pi* ‘this house/these houses’ would be grammatical, but **dak pi* and **da pik* (with the intended meaning ‘these houses’) would not.
- Although possible, languages with such rules seem to be somewhat counter-intuitive: if number is not an inherent property of a noun, as a lexical item, how can it spread to other constituents of the NP? Thus, 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. Hence, the feature of number cannot spread from the noun to the other elements of the NP, if the noun itself is underspecified for number.
- But however logical it may sound – almost to the point of appearing trivial – this claim certainly requires further reflexion and research.

References



- Corbett, Greville 2000. *Number*, Cambridge: Cambridge University Press.
- Corbett, Greville 2006. *Agreement*, Cambridge: Cambridge University Press.
- De Araujo, Gabriel Antunes 2004. *A Grammar of Sabané, a Nambikwaran Language*, Utrecht: LOT.
- 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.
- Fedden, Sebastian and Corbett, Greville 2017. “Gender and Classifiers in Concurrent Systems: Refining the Typology of Nominal Classification”, *Glossa* 2(1): 1-47.
- 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.
- Nordlinger, Rachel 1998. *A Grammar of Wambaya, Northern Territory (Australia)*, Canberra: Australian National University.
- 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.
- Smirnova, M. A. 1982. *The Hausa Language*, Moscow: Nauka.
- Smyth, David 2002. *Thai. An Essential Grammar*, London: Routledge.
- 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.
- Zubiri, Ilari 2000. *Gramática didáctica del euskera*, Bilbo: Didaktiker, S.A.
-

Thank you for your attention!



- E-mail: rmatasov@ffzg.hr
- Homepage: www.ffzg.hr/~rmatasov