Kaius Sinnemäki

A typological perspective on Differential Object Marking

Abstract: Case marking of the object is often claimed to be driven specifically by two referential properties of the object, animacy and definiteness. Data from 744 languages, however, provide typological evidence that there is no universal preference for object case marking to be driven by these properties, but at the same time provide strong evidence that object case marking tends to be restricted in some way rather than be generalized across all objects. I argue that the independence of object case marking from these two semantic-pragmatic properties may be explained by the instability of their relationship, and that economy provides a feasible explanation for restricting case marking to only some objects.

Keywords: Animacy, areality, case marking, definiteness, Differential Object Marking, economy, stability, typology

1 Introduction

Typological work often begins by comparing a few languages of different types, most often for purposes of updating typological theory or for describing linguistic diversity (Nichols 2007). Larger samples are commonly used only at a later stage to test hypotheses against a balanced sample and to assess the effect of different factors on the variable(s) of interest. Not unusually, though, hypotheses raised by research on small language samples have not been validated when using larger samples. For instance, many word order correlations previously thought to be universal have been shown to be mere areal tendencies (Bickel 2007).

Differential case marking of the object is an example of a grammatical pattern that has been intensely studied during the past 35 years but largely from a qualitative perspective. It has been studied in different languages (e.g., von Heusinger and Kornfilt 2005; Gerner 2007) and language families (e.g., Bossong 1991; Dimmendaal 2010) as well as from different theoretical viewpoints (e.g., Silverstein 1976; Comrie 1989; Enç 1991; Aissen 2003; Næss 2003). However, its typological distribution and properties have hardly ever been studied with large samples and
using quantitative methods (Bickel and Witzlack-Makarevitch [2008] is probably the first attempt towards this).

Differential case marking of the object (abbreviated as DOM) refers to the case marking of only a subset of objects by overt forms, often affected by the referential or discourse properties of the object. For instance, Turkish case-marks specific objects (1a), but leaves non-specific objects unmarked (1b).

(1) a. Ben bir kitab-ı oku-du-m.
   I a book-acc read-pst-1sg
   ‘I read a certain book.’
   (von Heusinger and Kornfilt 2005: 5)
b. Ben bir kitap oku-du-m.
   I a book read-pst-1sg
   ‘I read a book.’
   (von Heusinger and Kornfilt 2005: 5)

DOM has attracted a great deal of attention in linguistics, mostly due to its (allegedly uniform) crosslinguistic dependence on the referential properties of the object, most notably animacy and definiteness (or more broadly, on referential scales such as animacy hierarchy). Aissen (1999: 673), for one, claims that the dependence of DOM on such factors is among “the most robust generalizations in syntactic markedness”.

My purpose in this paper is to discuss the typological distribution of object case marking from the viewpoint of quantitative typology. In this approach, potential universals are best characterized as statistical generalizations, the probabilistic outcome of different forces at work, rather than as absolute inviolable linguistic constraints (e.g., Maslova 2009). Two issues are of particular interest here. First, I consider whether animacy or definiteness have a universal effect on the distribution of object case marking; these two properties are often thought to be the most important factors behind case marking (e.g., Aissen 2003: 436). Animacy and definiteness are commonly represented through hierarchies (Croft [2003: 130–132]; also Comrie [1989: 134–136]):

– Animacy: human < animate < inanimate.
– Definiteness: definite < specific < non-specific.

1 Abbreviations in interlinear glossing follow the List of Standard Abbreviations of the Leipzig Glossing Rules (http://www.eva.mpg.de/lingua/resources/glossing-rules.php) except for the following: AOR aorist, CLIT clitic, VAL validator.
As for object case marking, these hierarchies embody the generalization that the higher the object referent is on these hierarchies, the more likely it will be to be overtly case-marked. Instead of assessing the effect of such scales on case marking, as done by Bickel and Witzlack-Makarevitch (2008), I abstract away from these hierarchies and simply study whether animacy or definiteness as such have any relevance for the distributions of case marking worldwide. In other words, my aim is to study whether the occurrence of overt case marking is affected by these two properties, not whether it is the object referent higher on a particular hierarchy that tends to be marked.

Second, I abstract even further away from the properties that may affect variation in object marking and assess whether differential (or more broadly restricted) case marking is universally preferred over non-restricted case marking. Non-restricted case marking of the object refers here to overt case marking that is not limited to a subset of objects, but extends to all objects regardless of their properties. For instance, Imbabura Quechua uses the accusative case -ta for all kinds of objects, including definite human nouns (2a) and inanimate mass nouns (2b).

(2) a. Juzi Marya-ta juya-n-mi.
José María-ACC love-3-VAL
‘José loves María.’
(Cole 1985: 98)

b. Ñuka-ka aycha-ta miku-ni ñuka wawki tanda-ta miku-j-shna.
1-TOP meat-ACC eat-1 1 brother bread-ACC eat-NMLZ-as
‘I eat as much meat as my brother eats bread.’
(Cole 1985: 66)

The rationale for contrasting restricted case marking with non-restricted case marking is as follows. If semantic-pragmatic properties have no universal effect on the distribution of object case marking (as implied by the work of Bickel and Witzlack-Makarevitch [2008]), it may be that restricted case marking is still universally preferred over non-restricted case marking. Two issues suggest that this may indeed be the case. First, if the overt markers that express restricted (or differential marking) are lost, new markers often emerge to replace them, as has happened in Romance and Semitic (Bossong 1991: 152–154) and in Uralic languages (Havas 2008). Second, while the variation in object case marking appears to be affected by rather similar properties across languages, these properties may vary over time within a language. This has been documented in Persian (Hopper and Traugott [1993: 158–160], citing Bossong [1985]) and Romanian (von Heusinger and Onea Gáspár 2008). It is thus possible that neither the properties affecting the
language-internal distribution of overt case marking nor the formal case markers themselves are stable over time, but that the alleged stability would rather lie in the underlying system that is restricted regardless of changes in the formal or functional dimensions of case marking or in their synchronic manifestations at a given time. In fact, the little that has been reported about the distribution of these two types of case marking is that restricted (or differential) case marking is much more frequent and stable than non-restricted case marking (Bossong 1991: 152–154; Blake 2001: 120; Jäger 2007: 102).

According to data from a worldwide sample of 744 languages (see Section 2.3 for details of the sample), there was no typological evidence that animacy and/or definiteness has a universal effect on the distribution of case marking. However, the data provided statistical evidence that restricted case marking of the object was preferred over non-restricted marking. What these results suggest is that languages tend to restrict object case marking in some way but the details of this variation vary across languages more than has so far been assumed.

The rest of the paper is organized as follows. Section 2 presents the method and Section 3 the results, followed by discussion in Section 4 and conclusion in Section 5.

## 2 Method and data

### 2.1 Definitions

Differential Object Marking was defined above as overt case marking that is limited to a subset of objects and that is affected, for instance, by the referential or discourse properties of the object. It is common to delimit DOM to marking that depends *solely* on the semantic-pragmatic properties of the object, as in (1) (e.g., Aissen 2003). At other times a more inclusive approach has been taken, also subsuming the grammatical properties of the verb under the influencing factors (Malchukov and de Hoop 2010). For instance, Kashmiri case-marks (some) objects with the dative case in the non-perfective tenses (3a) but leaves the object unmarked in the perfective tenses (3b) (other restrictions apply; see Wali and Koul [1997]). This more inclusive latter definition was followed here to provide as broad coverage of object marking as possible. As a result, I classified object case marking as differential if it was affected by animacy, definiteness, information structure, kinship terms, proper/common distinction, or tense/aspect.
Note that DOM as defined here has a narrower scope than restricted marking. Restricted marking refers here to overt case marking that is limited to a subset of objects regardless of the influencing factors. For instance, plural objects are unmarked in Polish (4a) except for masculine referents, which are marked with the genitive case (4b) (other restrictions apply; Bielec [1998]). In Idoma, on the other hand, the object is unmarked in SVO word order (5a) but case-marked in SOV word order (5b). Case marking that depends on factors such as gender, number, or word order is not usually classified under DOM (see above) but comes naturally under restricted object marking. Therefore, case marking of the object was analyzed as restricted marking on a formal basis, glossing over the influence of different semantic-pragmatic properties on case marking.

(4) a. Kupilem drogie ołówki.
   buy.PST.1SG dear.NOM pencil.M.PL.NOM
   ‘I bought dear expensive pencils.’
   (Bielec 1998: 105)

b. Znam tych kolegów.
   know.1SG these friend.M.PL GEN
   ‘I know these friends.’
   (Bielec 1998: 106)

(5) a. Ô mà ɔyi.
   she bear child
   ‘She bore a child.’
   (Reh 1986: 121)

Note that animacy plays a role in the Polish object marking system (also in Russian and Belorussian, but the details vary): masculine human nouns are marked with the genitive case but masculine inanimate nouns are marked with the nominative case. However, animacy plays its role within the gender system and could be analyzed as a subclass of gender (e.g., Wiese 2011). For this reason I treat animacy under the gender system in Polish and classify its object case marking as restricted but not differential (similarly for Belorussian and Russian).
b. Ō  l-ŷì  mà.
   she  obj-child  bore
   ‘She bore a child.’
  (Reh 1986: 121)

As for non-restricted case marking of the object, it was defined in Section 1 as overt case marking that is not exclusively confined to a subset of objects, but extends to all objects. The “primary” type of non-restricted marking is what could be termed “generalized” case marking, whereby all objects receive the same case marker, as in Imbambura Quechua (2). In some languages, all objects are overtly case-marked but with different overt forms. Such splits in case marking are sometimes called symmetric differential case marking (de Hoop and Malchukov 2007). For instance, Georgian uses dative case to mark objects in the present tense (6a) but nominative case to mark objects in the past tense (6b). Regardless of such possible splits in object case marking, overt case marking extending to all objects was classified as non-restricted here.

(6) a. Glex-i  tesavs  simind-s.
   peasant-NOM  sow.3SBJ.3OBJ.PRS  corn-DAT
   ‘The peasant is sowing/sows the corn.’
   (Harris 1981: 1)

b. Glex-ma  datesa  simind-i.
   peasant-ERG  sow.3SBJ.3OBJ.AOR  corn-NOM
   ‘The peasant sowed the corn.’
   (Harris 1981: 1)

The role of animacy and definiteness in object marking was analyzed as follows. Case marking was classified as affected by animacy if animacy of the object referent itself played a role in case marking. Sometimes it was the object higher in animacy (e.g., human noun) and sometimes the object lower in animacy (e.g., inanimate noun) that was case-marked but this variation was ignored here (similarly for definiteness). In addition, case marking in many languages is used only for disambiguating who is doing what to whom (e.g., Yongren Lolo; Gerner [2007]). In these situations animacy of the object referent often plays a role (de Swart 2007), although other factors may be relevant as well (properties of other clausal participants, world knowledge). Such languages were here analyzed as being affected by animacy as well. Case marking was classified as affected by definiteness if it was affected by either definiteness or specificity, since these are often treated together (see Comrie [1989]; Croft [2003]). I further analyzed whether case marking was affected by one or both properties, animacy and/or definiteness.
(henceforth animacy/definiteness). While these two properties are logically independent of one another (as one anonymous reviewer pointed out), they often cluster in case marking as well as in other grammatical domains (e.g., Croft 2003: 166–175). For example, Spanish generally case-marks only human specific objects (von Heusinger 2008). The reason for the interaction between animacy and definiteness is only natural, given that they both are related to the individuation of an object (Hopper and Thompson 1980: 253).

For the purpose of typological comparison, object was more precisely defined as the more patient-like argument of a prototypical two-place transitive predicate (Comrie 2011). This definition of the object is not language-specific and enables crosslinguistic comparison in a feasible way (Haspelmath 2011). Object was deemed to be overtly case-marked when its marking deviated from that of the subject of an intransitive clause (Jäger 2003). For instance, if the absolutive argument was overtly case-marked, as in West Greenlandic (Fortescue 1984), it was not counted as object case marking.

A few constraints are appropriate to ensure greater comparability of case across languages. First, case marking was studied in affirmative transitive declarative active main clauses, whose object was a full noun phrase. Although case marking may vary according to polarity, mood, or clause type (e.g., Harris and Campbell 1995: 243; Aikhenvald and Dixon 1998), these were excluded to keep the workload manageable when working with a large sample. The exclusion of pronouns was justified by their high frequency in discourse, which increases structural conservativeness and makes grammatical structures more resistant to possible universal effects (Bybee and Thompson 2000). A good example is English, which has lost case marking of lexical objects but not of pronouns (*he hit the wall*/*him*). As a result, a possible universal preference for DOM will be more likely to be detected when focusing solely on noun objects (cf. Sinnemäki 2010: 872).

Second, case marking was defined broadly to cover all dependent marking of objects, following Nichols (1992). Variation in the formal realization of dependent marking was excluded, regardless of it being done by affixes, clitics, adpositions, particles, or by tonal or morphophonological alternations.

Third, possessed and lexically modified objects were excluded, since objects modified by adjectives, quantifiers, subordinate clauses, or possessives may behave differently from unmodified ones or those modified grammatically (e.g., by articles). For example, Hungarian uses accusative case only optionally when objects are marked with first or second person singular possessive suffixes, as in (7). Had such examples been included we would have missed the general trend in Hungarian to use accusative case for all unmodified objects regardless of their semantic properties.
(7) Látom a ház-am(-at).
   see:1SG the house-1SG(-ACC).
   ‘I see my house.’
   (Moravcsik 2003: 154)

Fourth, case marking of dislocated objects was excluded. Dislocation means that objects were set apart from the rest of the clause by a pause and/or replaced by a pronoun in situ. The rationale here is that dislocated objects are pragmatically marked as highly topical and may show unusual morphosyntactic coding compared to non-dislocated objects (see Iemmolo [2010]). On the one hand, they are often unmarked as opposed to case-marked non-dislocated objects, as in many Germanic languages (Boeckx and Grohmann 2004). On the other hand, they can be case-marked as opposed to unmarked non-dislocated objects. In Northern Italian (8) left-dislocated pronouns are marked with the preposition a ‘to’ as opposed to non-dislocated ones:

(8) A me, mi hanno chiamato subito.
   ACC me CLIT.1SG AUX.3PL call:PTCP:PST immediately
   ‘They called me immediately.’
   (Iemmolo 2010: 249)

These four criteria delimit object case marking to dependent-marked non-dislocated simple noun objects in affirmative transitive declarative active main clauses. The reasons for these constraints were both practical (to keep the workload manageable), and theoretical-methodological (to ensure the homogeneity of data and to focus on objects most prone to universal effects). However, because case marking can vary according to any of the criteria used, the extension of the study beyond this set of clauses would have inevitably corroborated the results (at least concerning the distribution of restricted vs. non-restricted case marking) and thus left the conclusions unaffected.

The hypotheses studied in this paper can be summarized as follows:

– Hypothesis 1: Distribution of object case marking is universally affected by animacy, definiteness, or animacy/definiteness.

– Hypothesis 2: Restricted case marking of the object is universally preferred over non-restricted case marking.

The null hypotheses are that the distribution of case marking is not affected by animacy or definiteness (Hypothesis 1) and that restricted marking is not universally preferred over non-restricted marking (Hypothesis 2). In the following section I outline the way in which these hypotheses were tested statistically.
2.2 Statistical testing of language universals

The typological distributions of object case marking were tested statistically to evaluate whether one pattern was universally preferred over another. In typology, a preference is normally deemed universal if two criteria are met (e.g., Bickel 2008a, 2008b): i) the preference is statistically significant and ii) it is independent of areal and genealogical factors.

To address the first criterion, the distribution of the case variables was tested with chi-square goodness-of-fit test against the null hypothesis of equal frequency (1:1 distribution). Since the sample was not drawn randomly and since many cell counts were small in a number of ensuing tests, the p-values were deduced via a randomized Monte Carlo permutation test (using 10,000 permutations), which is not vulnerable to small cell counts unlike classical statistical tests (cf. Janssen et al. 2007). Computation and plotting were done in the open-source statistical environment R (R Development Core Team 2011) and with relevant packages (vcd [Meyer et al. 2006, 2011]).

The second criterion is much more difficult to come to terms with. All statistical testing of universals has to deal with the confounding effects of borrowing (areality) and inheritance (genealogical relatedness), but there is no generally accepted way of doing this in typology. In the following, I outline the way these were dealt with in this paper.

The classical way to deal with the problem of inheritance is to compile a stratified random sample of mutually independent languages (e.g., Perkins 1989). However, stratified random sampling is difficult if not impossible in typology: on the one hand, sampling at the lowest taxonomic levels may inflate areal effects, and on the other hand, stratification is impossible in the case of isolates (Bickel 2008a: 222; Janssen et al. 2007: 420–424). Even if random sampling were possible, it may be simply unwanted, because a typologist generally wants to sample only well-described languages. Such principles of typological sampling are in direct conflict with the principles of (stratified) random sampling. In any case, the interpretation of the results from a stratified random sample would necessarily be synchronic in nature, providing static snapshots of currently spoken languages.

Dryer (1992) suggests addressing these problems by classifying languages into roughly comparable genealogical groups called genera and controlling how data points are counted in them. A genus is a genealogical group of languages that corresponds roughly to the branches of Indo-European (e.g., Germanic). The data points (or g-units) are then distinct values in genera. As a result, the most obvious source of genealogical bias is controlled by counting closely related languages with the same value only once.
A common problem with such sampling methods is that they are not very sensitive to diversity within families and tend to assume that all typological variables are equally stable within and across families (Bickel 2008a). However, the stability of typological variables depends on multiple competing factors and may vary both within and across families (Nichols 2003; Bickel 2013). It may thus be insufficient to stratify the sample by choosing one language per g-unit or by counting distinct values in g-units, because such sampling may underrepresent variation within the sampling unit. On the other hand, counting genera may overestimate diversity from the perspective of higher taxonomic levels, because closely related g-units are counted even when they behave uniformly (Bickel 2008a: 226). To better capture family-internal variation, sampling must be done not only across sampling units but also within them.

In order to deal with the problem of inheritance, I used controlled genealogical sampling (or g-sampling) as developed by Dryer (1992) and refined by Bickel (2008a). This method includes isolates and small families and therefore enables synchronic inferences from the distributions. Dynamic inferences can be drawn, for instance, by observing family-internal distributions in large-enough families (Bickel 2013).

G-sampling works in the following way. Let us first assume three families, A, B, and C, each consisting of ten sampled languages but with different distributions of the variable of interest, say, restricted and non-restricted marking (Table 1).

If the family behaved uniformly, it provided a single data point (family A; Table 1). Regardless of its cause (e.g., shared retention, innovation or family-bound drift; [Bickel 2008a: 225]) this bias needs to be controlled for, which was done by counting just one data point for the majority value. In case of diversity, the distribution was tested by chi-square goodness-of-fit test using randomized Monte Carlo permutation. If less than 5% of the simulations had a greater deviation than the observed data, the majority value was counted as a data point and the lan-

| Table 1: Possible distributions of restricted and non-restricted marking within families. |
|--------------------------------|---------|---------|---------|
| Restricted marking              | Family A | Family B | Family C |
|                                 |         |         |         |
| Non-restricted marking          |         |         |         |

3 G-sampling was computed by an R function developed by Taras Zakharko for Bickel (2008a). The function is available at http://www.spw.uzh.ch/software/gsample3.R (accessed 13 April 2012).
languages representing the deviating value were each counted as data points unless they dominated in a significant way at a lower taxonomic level. In family B, the observed distribution (1:9) deviated from the expected (5:5) in a significant way ($\chi^2 = 6.4, p = 0.02$), showing a bias to the majority value (non-restricted marking). In this family, languages with the majority value were treated as a homogeneous subgroup providing altogether one data point; the one deviating language must be due to non-genealogical factors and so it also provided a data point (Bickel 2008a: 225). In family C the observed distribution (6:4) did not deviate from the expected significantly ($\chi^2 = 2.7, p = 0.22$), and therefore the distribution was unlikely to have been genealogically induced. In these instances each sampled family member provided a data point: 6 restricted and 4 non-restricted marking data points (cf. Bickel 2008a: 226). The distribution of these data points was then assessed statistically.

The confounding effect of areality was assessed by evaluating the distributions of the variable of interest over geographic areas. It is important to test the effect of areality for (at least) two reasons. First, if languages did not behave similarly across geographic areas, even a biased distribution, such as 30\%:70\%, might be due solely to local historic accidents and not universal effects (see Perkins [1989]; Dryer [1992]; Bickel [2008b, 2013]). Second, earlier research suggests that the distribution of object case marking is areally biased (Bickel and Nichols 2009). These issues suggest that it is important to pay close attention to the areal aspects of object case marking.

Since it is unclear at what level of granularity areality would have an effect on object case marking, different areal resolutions were used (cf. Bickel 2008b). The most coarse-grained resolution was Nichols’ (1992) three-way areal breakdown into the Old World, the Pacific, and the New World. A slightly more fine-grained breakdown was Dryer’s (1992) six-way breakdown into Africa, Eurasia (except for mainland Southeast Asia), Southeast Asia-Oceania, Australia-New Guinea, North America (including Meso-America), and South America. This resolution was also chosen as the baseline, which is used in reporting the results in Section 3. The third and fourth resolutions were adopted from the Autotyp database (Nichols and Bickel 2009), the former dividing the world into ten (sub)continents (e.g., South and Southeast Asia, Australia) and the latter dividing the world into 24 smaller areas (e.g., California, North Africa, and Southeast Asia). The effect of areas was then tested with the chi-square test for independence.

---

4 See Nichols and Bickel (2009) for details. These areal breakdowns can also be visualized on a map by Google Earth: the files are available on the Autotyp website (http://www.spw.uzh.ch/autotyp/available.html, accessed 18 April 2012).
2.3 Sampling and data

In sampling, my only guiding principle was to sample as many genera as possible and to survey each area as thoroughly as possible. Aiming at maximum representativeness was motivated by two issues. First, to identify family-internal diversity, it was necessary to sample many languages from each family. Second, since the data was not sampled randomly, statistical inferences from sample to population were not warranted. Logical inferences are still possible, but only as far as the sample is representative across families as well (Janssen et al. 2007: 430). Hence it was meaningful to sample genealogical variation in each family and area as thoroughly as possible.

The resulting sample contained 744 languages from 389 genera (see Appendix). The data were drawn mostly from reference grammars (e.g., Cole 1985), descriptions of language families (e.g., Thurgood and LaPolla 2003), and works dealing with case marking or DOM (e.g., Bossong 1991). Languages were genealogically classified following the World Atlas of Language Structures (henceforth WALS; see Dryer and Haspelmath [2011]). Genealogical representativeness was evaluated against the total number of genera in the WALS in each baseline area. Overall it was over 75%, while area-wise it varied between 63% in South America and 98% in Eurasia (Table 2). As roughly 2/3 or more genera were sampled from each baseline area, the sample should be reasonably representative for reliable inferences.

Table 2: Genealogical coverage and areal distribution of the sample languages.

<table>
<thead>
<tr>
<th>Genera (sample/total)</th>
<th>Africa</th>
<th>Eurasia</th>
<th>SEA–Oc.</th>
<th>Aus.–NG</th>
<th>N Am.</th>
<th>S Am.</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Genera</td>
<td>69/73</td>
<td>42/43</td>
<td>49/63</td>
<td>93/139</td>
<td>76/94</td>
<td>60/96</td>
<td>389/508</td>
</tr>
<tr>
<td>Genealogical coverage</td>
<td>95%</td>
<td>98%</td>
<td>78%</td>
<td>67%</td>
<td>81%</td>
<td>63%</td>
<td>77%</td>
</tr>
<tr>
<td>Languages</td>
<td>164</td>
<td>95</td>
<td>117</td>
<td>135</td>
<td>134</td>
<td>99</td>
<td>744</td>
</tr>
</tbody>
</table>

A few genera not represented in the latest edition of the WALS (Dryer and Haspelmath 2011) were added (e.g., Tujia in Tibeto-Burman).
3 Results

3.1 Descriptive results

There were 223 languages (30%) with object case marking in the sample and 521 (70%) without it. Restricted case marking occurred in 178 languages (80% of languages with object case marking), while non-restricted case marking occurred in 45 languages (20%). Variation in object case marking was affected by animacy in 52 languages (29% of languages with restricted case marking of the object), by definiteness in 49 languages (28%), by both properties in 22 languages (12%), and by some other property in 55 languages (31%). Note that definiteness influenced object case marking equally often than did animacy.

Fig. 1 presents the distribution of object case-marking languages of different types on a world map. The distribution of restricted and non-restricted case marking of the object does not appear to be particularly biased areally, owing to the geographically relatively even distribution of non-restricted marking. As for the object properties, definiteness seems to affect case variation especially in the Old World but less so elsewhere. Other conclusions are difficult to draw on the basis of visual patterns alone.

The distribution of g-units with or without object case marking was not very far from 1:1 (44% and 56% respectively). However, the share of object-marking g-units was unexpectedly high. Earlier research on case alignment has shown that accusative alignment occurs in only about 25–30% of data points (Nichols 1992: 90; Comrie 2011). This discrepancy compared to the present study stems from a methodological issue. Namely, Bickel's (2008a) g-sampling captures diversity better than do stratified sampling or counting genera (cf. Dryer 1992). This is evident from the somewhat high share of object-marking g-units (44%) compared to the much lower share of object marking when counting languages (30%) or genera (31%). What this issue illustrates is that methodological choices can have a real effect on quantitative results.

3.2 The effect of referential properties

When testing Hypothesis 1, I assessed whether animacy or definiteness had an effect on the distribution of restricted object marking. The areal distributions of g-units that were vs. those that were not affected by these properties are presented in Fig. 2–4. The results of the statistical tests are summarized in Table 3 for the baseline areal breakdown. In the other areal breakdowns the results were
Fig. 1: The case-marking sample languages with non-restricted marking (black circles), and restricted marking influenced by animacy (white squares), definiteness (white triangles), both (white diamonds), and neither (white circles) (map produced by the map-generating tool of the WALS, developed by Hans-Jörg Bibiko).
very similar and their statistics are therefore reported only if they differed from those of the baseline areal breakdown.

As for animacy, the number of g-units not affected by animacy was higher than that of g-units affected by animacy in all areas except in North and South America (see Fig. 2). The chi-square statistic in the test for independence was small and far from statistical significance ($\chi^2 = 4.2; p = 0.53$; Table 3): the effect of animacy on case marking was thus fairly similar across areas. Animacy had an effect on object case marking in 47 g-units but no effect in 75 g-units and this distribution was statistically significant ($\chi^2 = 6.4; p = 0.013$; Table 3). This result means that there is a crosslinguistic dispreference for object case marking to be driven by animacy.

The results for definiteness were more complex. In all areas except Africa there were fewer g-units in which definiteness had an effect on object case marking than those in which it had no effect (Fig. 3). In line with this, the chi-square statistic in the test for independence was significant ($\chi^2 = 20.9; p = 0.0009$), meaning

Table 3: Statistics for the effect of animacy and definiteness on object case marking.

<table>
<thead>
<tr>
<th>Object property</th>
<th>Effect of areas</th>
<th>Effect of object properties</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\chi^2$</td>
<td>$p$</td>
</tr>
<tr>
<td>Animacy</td>
<td>4.2</td>
<td>0.53</td>
</tr>
<tr>
<td>Definiteness</td>
<td>20.9</td>
<td>0.0009</td>
</tr>
<tr>
<td>Animacy/definiteness</td>
<td>3.7</td>
<td>0.60</td>
</tr>
</tbody>
</table>
that the effect of definiteness on object case marking was not similar across areas. In other words, although definiteness was significantly dispreferred (43 g-units affected by definiteness and 83 g-units not affected by it; $\chi^2 = 15.0; p = 0.0002; \text{Table 3}$), this dispreference was local: about 78% of g-units affected by definiteness were in the Old World and only about 22% elsewhere. Thus, while object case marking seems to be affected by animacy rather similarly across areas, it is affected by definiteness mostly in the Old World.

I now evaluate whether one or both of the two properties, animacy and definiteness, had an effect on object case marking. Animacy/definiteness seemed to have an effect on object case marking in rather similar ways across areas (Fig. 4). In line with this the chi-square statistic in the test for independence was small and non-significant ($\chi^2 = 3.7; p = 0.60; \text{Table 3}$). The distribution of case-marking g-units with or without an effect from animacy/definiteness (70 vs. 51 g-units respectively) was likewise non-significant ($\chi^2 = 2.98; p = 0.10; \text{Table 3}$). This result
means that there is no evidence of any universal effect of animacy/definiteness on object marking.

I then analyzed whether animacy or definiteness had any effect on case marking in DOM languages. According to the results, the effect of animacy was neither preferred nor dispreferred ($\chi^2 = 0.9; p = 0.39$), although it was independent of area ($\chi^2 = 6.9; p = 0.23$). The effect of definiteness depended again on area ($\chi^2 = 35.1; p < 0.0001$): 81% of g-units driven by definiteness were in the Old World and only 19% elsewhere. When studying whether one or both of these referential properties had an effect on object case marking, there was a significant tendency for DOM to be affected by animacy/definiteness ($\chi^2 = 27.8; p < 0.0001$), but this was not independent of area ($\chi^2 = 15.4; p = 0.0083$). These results mean that there is insufficient evidence to assume that the effect of animacy/definiteness on DOM is similar across areas.

All in all, there was no evidence for a systematic co-occurrence between case marking and the object properties scrutinized. These results were unexpected given the prominent role attributed to animacy and definiteness in the literature, calling for explanations, on the one hand, and further research, on the other.

### 3.3 Restricted vs. non-restricted case marking

The distribution of restricted and non-restricted object marking over the baseline areas is shown in Fig. 5. The frequency of restricted marking seems to exceed that of non-restricted marking in all areas. In line with this, the chi-square statistics in the test for independence were small and far from statistical significance, regardless of the areal breakdown ($p > 0.10$; the leftmost columns of Table 4). This suggests that the distribution of restricted and non-restricted marking was independent of area.

The distribution of these case marking patterns was biased to restricted marking. There were altogether 111 restricted-marking g-units (74%) and 40 non-restricted-marking g-units (26%), suggesting that neither variable is particularly rare among object-marking languages. These frequencies deviated from

---

6 The effect of animacy on DOM was not independent of areas in the 24-way areal breakdown ($\chi^2 = 33.0; p = 0.015$) or in the 3-way areal breakdown ($\chi^2 = 6.7; p = 0.034$). There is thus some evidence that the effect of animacy on DOM is geographically restricted.

7 The frequency of non-restricted object marking was 8.9% when compared to all patterns of object case marking (including zero-marking; counting genera). This roughly corresponds to the frequency of object-subject word order (6.2%; Dryer [2011]), which is generally considered cross-linguistically very rare (Comrie 1989).
the expected ones in a significant way, regardless of the areal breakdown ($\chi^2 > 32; p < 0.0001$; middle columns of Table 4). This suggests that restricted marking was significantly preferred over non-restricted marking and that this preference was independent of area. Synchronically speaking, a strong universal preference existed for restricted marking.

I also assessed the reliability of the results by adapting the reliability method of Janssen et al. (2007: 435–437). The crux of this test was to incrementally increase the number of events predicted to be dispreferred in the goodness-of-fit tests (here the non-restricted marking events) or alternatively to decrease the number of events predicted to be preferred (here the restricted marking events) until statistical significance was lost. If the significance was lost by adding, say, one or two dispreferred events, or by removing one or two preferred events, one should be very cautious in interpreting the results.
The rightmost columns of Table 4 present the results of these tests. The N-restr.-column shows the increase of non-restricted-marking g-units (they were dispreferred) that could have occurred for the result to have remained statistically significant (p < 0.05); the Restr.-column shows the corresponding tolerated decrease of restricted-marking g-units (they were preferred). This test simulated how resistant the results were to changes in the sample size. The Both column reports the results of a test simulating how well the result tolerated misclassifications. This simulation was done by simultaneously increasing the number of non-restricted marking units by one and decreasing the number of restricted marking units by one until the result was no longer statistically significant (p < 0.05). Since the result was resistant to 22 or more misclassifications or alternatively at least 42 changes (28%) in sample size to the predicted preference, our confidence in the results was greatly strengthened.

To conclude this section, we can state that the data provided solid evidence for Hypothesis 2. Namely, restricted marking was preferred over non-restricted marking in a statistically significant way, and importantly, regardless of both genealogical and areal factors. The overall results in Section 3 suggest that variation in object case marking is not universally conditioned by the referential properties typically associated with differential marking (that is, animacy and definiteness) but by formal factors related to the occurrence of object marking (restricted vs. non-restricted).

4 Discussion

The results reported in Section 3 raise three issues in particular: i) the absence of a preference in object-marking languages for case marking to be affected by animacy or definiteness, ii) the preference for restricted object marking over non-restricted marking, and iii) the existence in the first place of non-restricted object making (cf. Harris 2008: 54). These issues are discussed in the following sections.

4.1 Case marking and the effect of semantic-pragmatic properties

It has been a common belief among many linguists that object case marking is closely associated with the semantic-pragmatic properties of the object, especially animacy or definiteness. However, the results provided typological evi-
idence against a universal dependence between object case marking and these referential properties of the object.

I suggest that this outcome is part of a more general trend where dependence on semantic-pragmatic properties can increase the instability of a grammatical pattern. It seems a rather general property of grammatical patterns that if they are influenced by semantic-pragmatic properties, they may be rather unstable. Based on statistical measures of stability, Wichmann and Holman (2009) argue that among the WALS features the basic morphosyntactic features tend to be stable (e.g., word order, affix order), while features influenced by semantics or pragmatics (e.g., definite and indefinite articles, politeness distinctions in pronouns, imperative-hortative systems, and the coding of epistemic or evidentiality distinctions) tend to be unstable. A possible reason for this asymmetry is that morphosyntactic features, such as word order parameters, tend to correlate with one another (e.g., Dryer 1992), while the role of semantic-pragmatic properties is prone to vary locally (e.g., in terms of culture) and/or randomly (cf. Wichmann and Holman 2009).

Instability seems also to be characteristic of restricted case marking. There were 39 genera in my sample with two or more object-marking languages and among these there were only five genera (13%) in which case marking depended on the same set of semantic-pragmatic properties across languages (the total set of properties studied included animacy, definiteness, common/proper, kin terms, tense/aspect, information structure, and disambiguation). In other words, while case marking is typically dependent on some semantic-pragmatic property, there is great diversity in terms of which properties are involved in a given language. An additional complicating factor is that these properties may vary over time in a particular language, as reported in many languages (e.g., Romanian, von Heusinger and Onea Gáspár [2008]; Spanish, von Heusinger [2008]; and Persian, Hopper and Traugott [1993: 158–160]). As a result, while certain properties of argument marking may be stable (cf. Nichols 1992), other properties related to case marking may be quite unstable (also Mithun [2012]).

4.2 Frequency asymmetries in object case marking

As the results suggested, the frequency asymmetry between restricted and non-restricted case marking was very strong in the light of typological data. This asymmetry naturally raises the question why non-restricted marking exists at all. Two possible reasons are discussed first before discussing the preference for restricted marking.
First, non-restricted marking exists because it is a possible outcome of a regular grammaticalization mechanism, namely, semantic extension. When languages develop case marking of object, it usually begins with the most prominent objects and is used for argument discrimination or for indexing some object property (Comrie 1989; de Swart 2007; Iemmolo 2010). Via extension the occurrence of case marking may spread to more contexts gradually leading to the marking of less prominent objects. Finally, case marking may spread to all objects, reaching the final point of semantic extension (called total niche take-over by Dahl [2004: 128]). In this process, the indexing function of case marking simplifies conceptually from its earlier stages as the conditions for its use are dropped until semantic weakening reaches its endpoint, when restricted case marking turns into non-restricted case marking and there is nothing more to index but the objecthood of the argument (Hopper and Traugott [1993: 87–93], citing Bosson [1985]; Harris and Campbell [1995: 113–115]). This process illustrates that non-restricted case marking can be a natural outcome of a common grammaticalization mechanism.

However, we do not know enough about the historical changes in languages to be certain that non-restricted case marking develops in languages through semantic extension. The little that we know is that non-restricted case marking may not arise as slowly as suggested by the regular process of semantic extension. Consider, for example, case marking in Hungarian, Awngi, and Evenki.

While many Uralic languages have retained the Proto-Uralic Differential Object Marking, Hungarian eventually lost it and developed object case marking that is non-restricted (Havas 2008). Awngi, a Central Cushitic language with non-restricted case marking, has retained object case marking from the Proto-Central Cushitic, but has lost differential marking as well as the cognate case forms mostly shared by other Central Cushitic languages (Appleyard 1988). In other words, both Hungarian and Awngi have lost their old differential case marking systems and have developed non-restricted marking anew, and all the while many languages in their families have retained the differential case system and even the case forms from the respective Proto-language. These examples do not mean that non-restricted marking could develop directly without case marking passing through a period of restricted marking; however, they suggest at least that the period of restricted marking may not be very long and stable.

In addition, the other subtype of non-restricted marking, namely split case marking, may develop in a similar way. Evenki is a Tungusic language which has non-restricted case marking. Accusative case occurs in all Tungusic languages and the case forms are cognate in all of them (-vA/be; see Pakendorf [2007:

---

8 I am grateful to an anonymous reviewer for helpful comments on Hungarian.
However, Evenki (and apparently Negidal) have developed a separate marker for indefinite objects, namely -(y)A. Thus while other Tungusic languages have retained their Differential Object Marking, Evenki has developed non-restricted marking, not through semantic extension but by employing a separate case form for indefinite objects.

The second reason why non-restricted marking exists is that it is functionally motivated, by transparent form-meaning mapping. On the one hand, when case marking is generalized to all objects, overt case marking is affected only by the objecthood of the argument, and contextual cues are not relied on. On the other hand, when overt case marking is split into different forms, as in Evenki, the form-meaning mapping is more transparent than in the case of restricted case marking, which has to use different contextual cues to disambiguate the zero-marked arguments.

However, while transparency may sometimes facilitate language processing, the use of contextual cues may be more efficient than overt coding from the viewpoint of language processing. What is relevant in terms of processing is that contextual cues seem to be used anyway and if objecthood can be reliably assigned by virtue of such cues, identification of objecthood becomes inefficient if objects are case-marked nevertheless, since overt marking increases structural complexity and requires the processing of additional forms (Hawkins 2004: 48). Although transparent form-meaning mapping may motivate non-restricted marking, it requires more morphemes and thus greater structural complexity than required by restricted case marking, which leaves some objects unmarked. Structural transparency may thus not be a strong motivation for case marking, and this may partly explain the overall rarity of non-restricted marking.

This last point leads to the last question discussed in this connection, namely the universal preference for restricted instead of non-restricted case marking of the object. Because Differential Object Marking and the reasons underlying it have been discussed so extensively in the literature, I will only briefly raise what I perceive to be the main reasons behind the preference for restricted over non-restricted case marking.

First, while grammaticalization as a diachronic process is mostly unidirectional, the spread of case marking from restricted to non-restricted is reversible due to the decay of case marking (Bossong 1991: 152; König 2008: 288). This means that when non-restricted case marking decays, the process is not abrupt but progressive, leading to zero via restricted marking. For example, Old Harari (South Semitic) apparently had non-restricted case marking, marking both definite and indefinite objects nouns with accusative case, but in Modern Harari only definite objects are marked owing to case decay (König [2008: 226–227], citing Tosco [1994]). Many other Semitic languages (e.g., Aramaic and Amharic) have
also replaced the Proto-Semitic non-restricted object marking by restricted marking (Bossong 1991: 145–146, 149–151). Restricted marking can thus develop when new case marking emerges in a language but also when old non-restricted marking decays. This means that there are more possibilities for restricted marking to develop, which results in greater probability of change therein.

Second, case marking may be blocked from extending to all objects because the absence of overt marking is reserved for some other grammatical function. This other function has to do with the tightening of the relationship between object and verb (Dahl 2004: 253, 2009: 245). Low-referentiality nouns are especially prone to so-called quasi-incorporation, in which elements of a construction are closely knit but not fully incorporated, resembling prototypical incorporation but without the incorporated element losing its word status (Dahl 2004: 216–219). Quasi-incorporation makes the object indefinite and non-referential and is typically signaled by reduced grammatical marking of the object noun, such as the loss of number, case marking, modifiers, and/or determiners. In Hungarian, quasi-incorporation is signaled by dropping plural marking from the noun: if dropped (9b), both singular and plural interpretations are allowed, but if retained (9a), plural reference is clearly intended. In Turkish, quasi-incorporation is signaled by dropping case marking and the article: compare (10) with (1a) and (1b).

(9) a. Őva level-ek-et ír.
    Őva letter-PL-ACC write.PRS.3SG
    ‘Éva is writing letters.’

b. Pisti level-et ír.
    Pisti letter-ACC write.PRS.3SG
    ‘Steve is writing letters/a letter (is engaged in letter-writing).’
    (Kiefer 1990: 151, via Dahl 2004: 216)

(10) Ben kitap oku-du-m.
    I book read-PST-1SG
    ‘I was book-reading.’
    (von Heusinger and Kornfilt 2005: 5)

Because the dropping of case marking of non-referential objects has this well-defined and well-motivated function in Turkish and many other languages, it can effectively block case marking from spreading to non-restricted marking and continue to restrict the object case marking to certain contexts only.

The third, and probably the most important reason for the strong preference for restricted marking is arguably the Zipfian economy effect of frequency of use.
It is well known that inanimate and/or indefinite nouns function most often as objects, while animate and/or definite nouns function as subject – by virtue of economy, only the less frequent objects in discourse, that is, animate and definite ones, are typically case-marked (Haspelmath 2008a: 15). However, the relevance of the frequency asymmetry to object case marking is not limited to animacy and definiteness but concerns other properties as well. If word order was the deciding factor behind case marking in the sample languages, object was generally case-marked in the less frequent word order \((n = 17; 89\%\); counting genera). If the decisive factor was information structure, it was generally the topical or emphasized object that was case-marked rather than the more frequent non-topical or non-emphasized object \((n = 17; 100\%)\). If the decisive factor was kinship terms or proper/common distinction, it was the kinship terms and proper nouns that were marked rather than the more frequent non-kinship terms or common nouns \((n = 10; 71\%)\).

What these points suggest is that variation in object case marking may be better explained by economy than by particular semantic-pragmatic properties or referential hierarchies (cf. Haspelmath 2008a; Bickel and Witzlack-Makarevitch 2008). This is not bad news for functional-typological linguistics, since formal economy, such as in restricted object marking, is supported by preferences in language processing (e.g., Hawkins 2004) and these are probably the most convincing functional explanations available (cf. Haspelmath 2008b).

### 5 Conclusion

The following conclusions can be drawn from the statistical analyses of object case marking in this paper. There was no typological evidence for object case marking being affected by animacy or definiteness of the object referent. This result was claimed to stem from the generally low stability of grammatical patterns influenced by semantic-pragmatic properties. However, there was strong typological evidence for restricted case marking of object to be universally preferred over non-restricted marking. The main reasons behind this result were argued to stem from the higher probability of change to restricted marking, its better functional motivation especially by economy, and the fact that case marking may be blocked from extending to non-restricted marking by the use of zero-marking of the object for some other grammatical function. All in all, while object case marking very often co-occurs with some properties of the object referent, there is no universal preference for object case marking to be affected by one or both of the two referential properties, animacy or definiteness, but different languages appear to grammaticalize in different ways.
Acknowledgments: I would like to thank Giorgio Iemmolo, Gerson Klumpp, Matti Miestamo, Fred Karlsson, Martin Haspelmath, and three anonymous referees for valuable comments that helped to improve the manuscript. I am also grateful for the financial support from The Emil Aaltonen foundation, which made this research possible.

References


Appendix: The sample languages

The following is a list of the sample languages by area and genus (in italics). Languages are in parentheses after each genus, unless the name of the genus and the single language representing that genus coincide. The classification of overt case marking of object is provided in square brackets after language name: n = non-restricted, r = restricted, a = animacy, d = definiteness.

Africa: =|Hoan [r], Adamawa (Mbum, Mumuye), Bantoid (Babungo, Kongo, Lunda, Luvale, Makua, Ndebele (in South Africa), Nkore-Kiga, Swahili, Tunen, Zulu), Beja [n], Berber (Berber (Middle Atlas), Kabyle), Berta, Biu-Mandara (Gude [rd], Hdi [r], Margi), Bongo-Bagirmi (Bagirmi), Central Cushitic (Awngi [n], Bilin [rd], Kemant [rd]), Central Khoisan (Khoekhoe [n]), Cross River (Efik, Kana, Lokaa, Obolo, Ogbonuagum), Daju (Shatt, Sila), Defoid (Yoruba), Dogon (Dogon [ra], Dogon (Toro Tegu), Jamsay, Tommo So [rad]), East Chadic (Kera, Lele), Eastern Cushitic (Oromo (Harar), Somali), Eastern Jebel (Inessana), Eastern Mande (Boko, Busa, Guro, Yaouré), Eastern Saharan (Kanuri [ra], Tubu [rd]), Edoid (Degema, Emai, Engenni), Egyptian-Coptic (Coptic [rd]), Fur [rad], Gumuz, Gur (Dagaare, Dagbani, Gurenne, Kasem, Koromfe, Ngangam, Sucite, Supyire), Hadza, Heiban (Moro), Idomoid (Idoma [r], Igede), Igboi (Igbo, Izi), Ijoid (Defaka, Ijo (Kolokuma)), Kadugli (Krongo), Kainji (ut-Ma’in [n]), Katla-Tima (Katla), Koman (Komo), Kresh (Aja, Kresh), Kru (Grebo, Klao), Kuliak (Ik [n], So), Kunama [r], Kwa (Akan (Asanti), Akan (Fante), Ewe, Kpoko), Laal, Lendu (Ngiti), Maban (Maba [rd], Masalit [r]), Mangbetu, Mangbutu-Efe (Mamvu), Masa (Masa, Musey), Moru-Ma’di (Avokaya, Lugbara, Ma’di), Nilotic (Bari, Dinka, Lango, Luo, Maasai,
Päri, Turkana), North Omotic (Basketo [n], Dizi [rd], Gamo), Northern Atlantic (Diola-Fogny, Fula (Cameroonian), Fula (Nigerian)), Northern Khoisan (Ju’hoan [r]), Nubian (Ghulfan [rd], Nobin [n], Nubian (Dongolese) [rad]), Nupoid (Gwari), Nyimang [n], Platoid (Biom, Fyem, Kagoma), Rashad (Orig), Sandawe, Semitic (Amharic [rd], Arabic (Egyptian), Harari [rd], Hebrew (Modern) [rd], Maltese, Tigré [rd], Tigrinya [rad]), Songhay (Koyra Chiini, Koyraboro Senni, Tadaksahak), South Omotic (Dime [n]), Southern Atlantic (Kisi), Southern Cushitic (Iraqw [n]), Southern Khoisan (!Xóõ, /Xam [r]), Surmic (Didinga, Murle, Mursi), Talodi Proper (Masakin), Taman (Mararit [n], Misiiri [n], Tama [rad]), Tegem (Lafafa), Temein, Ubangi (Gbaya Kara, Gbeya Bossangoa, Mbodomo, Ndogo, Ngbaka, Sango, Zande), West Chadic (Hausa, Miya), Western Mande (Bambara, Jowulu, Kpelle, Ligbi, Mandinka, Maninka (Western), Mende, Soninke, Susu, Vai, Yalunka).

Australia–New Guinea: Angan (Hamtai), Anindilyakwa, Annabel (Rao), Awin-Pare (Kamula), Awju-Dumut (Kombai, Wambon), Bilua, Binanderean (Suena), Border (Amanab [r], Imonda [ra]), Bosavi (Edolo), Bulaka River (Maklew), Bunuban (Gooniyandi), Burarran (Burarra), Chimbu (Kuman, Nii), Dagan (Daga), Dani (Dani (Lower Grand Valley)), East Bird’s Head (Meyah, Sough), East Geelvink Bay (Bauzi), East Strickland (Nomad), Eastern Highlands (Fore [ra], Gahuku [r], Hua), Engan (Kewa), Fasu [rd], Gapun (Taiap), Garrwan (Garrwa), Gogodala, Goilalan (Kunimaipa, Weri), Greater Alor (Adang [r], Woisika), Gunwinygic (Bininj Gun-Wok), Hatam, Iwaidjan (Maung), Jaminjung (Jaminjung), Karkar-Yuri, Kaure, Kebar (Mpur), Kiwaian (Kiawai), Koiarian (Barai, Koiari), Kolopom (Kimaghama), Kombio-Arapesh (Arapesh (Mountain)), Kuot, Kwerba (Isirawa [r]), Kwomtari (Kwomtari [n]), Lakes Plain (Iau, Kirikiri), Laragiyan (Laragia), Lavukaleve, Lower Ramu (Watam [ra]), Lower Sepik (Yimas), Madang (Amele, Kobon, Mugil, Nend, Siroi, Tauya, Usan), Mailuan (Magi), Mairasi, Makasae-Fataluko-Oirata (Makasae [r]), Mangarrayi [r], Maran (Alawa, Warndarang), Marind Proper (Marind, Yaqay), Mek (Una), Middle Sepik (Ambulas [n], Iatmul, Kwoma [r], Seim [n]), Morehead and Upper Maro Rivers (Kanum (Bädi)), Murrinh-Patha, Ngalakan, Nimboran, North Halmaheran (Sahu, Ternate, Tobelo), North-Central Bird’s Head (Abun, Maybrat), Northern Daly (Malakmalak), Nunggubuyu, Nyulnyulan (Warrwa), Ok (Telefol), Oksapmin, Orya [r], Pama-Nyungan (Diyari [r], Dyirbal, Gumbaynggir [r], Kalkatungu, Martuthunira [n], Ngarla, Ngiyamba, Nhanda [rd], Pitjantjatjara, Pitta Pitta [n], Wangkumara [n], Yidiny), Pawaian, Ram (Awetu [ra]), Senagi (Anggor, Dla (Proper) [r]), Sepik Hill (Alamblak), South Bird’s Head (Inanwatan), Southern Daly (Ngankikurungkurr), Sulka, Tama Sepik (Yessan-Mayo [ra]), Tangkic (Kayardild [n], Yukulta), Tate (Kaki Ae), Tiwian (Tivi), Tor (Berik [r]), Turama-Kikorian (Rumu), Urim, Wapei-Palei (Au, Olo, Walman), West Barkly (Djingili, Wambaya [n]), West Bougainville
(Rotokas), *West Wapei* (One), *Western Daly* (Maranungku, Marrithiyel), *Western Skou* (Dumo, Skou), *Wissel Lakes-Kemandoga* (Moni), *Wororan* (Ungarinjin), *Yale* (Nagatman [r]), *Yangmanic* (Wardaman), *Yawa* (Saweru), *Yele* (Yeli Dnye), *Yellow River* (Namia [rd]).

**Eurasia:** *Ainu*, *Albanian* [rd], *Armenian* (Armenian (Western)), *Avar-Andic-Tsezic* (Avar, Hunzib), *Baltic* (Latvian [r], Lithuanian [n]), *Basque*, *Burushaski*, *Celtic* (Breton, Cornish, Gaelic (Scots) [rd], Irish, Welsh), *Central Dravidian* (Kolami [ra]), *Finnic* (Estonian [r], Finnish [r]), *Germanic* (Afrikaans, Dutch, English, German [r], Icelandic [n], Norwegian, Swedish), *Greek* (Greek (Modern) [r]), *Indic* (Assamese [rad], Hindi [ra], Kashmiri [ra], Marathi [rad], Nepali [rad], Panjabi [rad], Rajbanshi [rad]), *Iranian* (Kurdish (Central), Ossetic [rd], Persian [rd]), *Japanese* [r], *Kartvelian* (Georgian [n]), *Korean* [ra], *Kusunda* [ra], *Lak-Dargwa* (Dargwa, Lak), *Lezgic* (Lezgian), *Mari* (Mari (Meadow) [n]), *Mongolic* (Buriat [rd], Khalkha [rd]), *Mordvin* (Erzya [rd]), *Munda* (Korku [ra], Mundari), *Nakh* (Chechen, Ingush), *Nivkh*, *Northern Chukotko-Kamchatkan* (Chukchi), *Northern Dravidian* (Brahui [rd]), *Northwest Caucasian* (Abkhaz, Adyghe (Abzakh), Kabardian), *Perm* (Komi-Zyrian [rad], Udmurt [rd]), *Romance* (French, Italian, Provençal, Sardinian [rad], Spanish [rad]), *Saami* (Saami (Central-South) [rd], Saami (Skolt) [r]), *Samoyedic* (Enets [rd], Kamass [rd], Nenets [rd], Selkup [rd]), *Slavic* (Belorussian [r], Macedonian, Polish [r], Russian [r]), *South-Central Dravidian* (Telugu [ra]), *Southern Chukotko-Kamchatkan* (Itelmen), *Southern Dravidian* (Kannada [rad], Kota [ra], Malayalam [ra]), *Tungusic* (Even [n], Evenki [n], Manchu [rd], Nanai [r], Orok [r], Udhe [rd]), *Turkic* (Azerbaijani [rd], Bashkir [rd], Tatar [rd], Turkish [rd], Tuvin [rd], Uzbek [rad]), *Ugric* (Hungarian [n], Khanty (Eastern), Mansi [rd]), *Yeniseian* (Ket), *Yukaghir* (Yukaghir (Kolyma) [n]).

**North America:** *Aleut*, *Algonquian* (Cree (Plains), Passamaquoddy-Maliseet), *Alsea*, *Amuzgoan* (Amuzgo), *Atakapa*, *Athapaskan* (Navajo, Slave, Tanacross), *Aztecan* (Nahuati (Huasteca), Nahuatl (Michoacán), Nahuatl (North Puebla), Nahuatl (Tetelcingo), Pipil), *Caddoan* (Wichita), *Cahita* (Yaqui [r]), *Central Salish* (Halkomelem, Squamish), *Chichimec* (Chichimeca-Jonaz), *Chimacuan* (Quileute [n]), *Chimariko*, *Chinantecan* (Chinantec (Comaltepec), Chinantec (Lalana), Chinantec (Lealao), Chinantec (Quiotepec), Chinantec (Sochiapan), Chinantec (Tepetotutla)), *Chinookan* (Chinook (Upper)), *Chitimacha*, *Chumash* (Chumash (Barbareño)), *Coahuiltecan* (Coahuitlaco), *Corachol* (Cora), *Costanoan* (Mutsun [n]), *Eskimo* (Greenlandic (West), Yup’ik (Central)), *Eyak*, *Haida*, *Hopis* [n], *Huavean* (Huave (San Mateo del Mar)), *Interior Salish* (Shuswap), *Kalapuyan* (Kalapuya), *Karok*, *Keresan* (Acoma), *Kiowa-Tanoan* (Kiowa, Tiwa (Northern)), *Klamath-Modom* (Klamath [ra]), *Kutenai*, *Mayan* (Chontal Maya, Huastec, Jakaltek,
Kanjobal (Western), K’ekchi, Sipakapense, Tzeltal (Petalcingo), Tzutujil, Yucatec, Misumalpan (Miskito, Sumu), Miwok (Miwok (Southern Sierra) [n]), Mixe-Zoque (Mixe (Isthmus), Ototec, Zoque (Chimalapa), Zoque (Copainalá)), Mixtecan (Mixtec (Alacatlatzala), Mixtec (Chalcatongo), Mixtec (Diuxi-Tilantongo), Mixtec (Jamiltpec) [ra], Mixtec (Jicaltepec), Mixtec (Ocotepec), Mixtec (Silacayoapan), Mixtec (Yosondúa)), Molala [ra], Muskogean (Choctaw [r], Koasati [ra]), Natchez, Northern Iroquoian (Mohawk, Oneida), Numic (Comanche [r], Paiute (Northern) [r], Ute [n]), Otomian (Ocuilteco, Otomí (Mezquital)), Pomoan (Pomo (Eastern) [r], Pomo (Northern) [ra], Pomo (Southeastern) [ra]), Popolocan (Mazatec (Chiquihuitlan), Mazatec (Huautla)), Sahaptian (Nez Perce [n]), Salinan, Seri, Shasta, Siouan (Lakota, Osage), Southern Iroquoian (Cherokee), Southern Wakashan (Makah, Nuuchahnulth), Subtiaba-Tlapanec (Tlapanec), Takelma, Takic (Cahuilla [r], Luiseño [ra]), Tarahumaran (Tarahumara (Western)), Tarascan (Purépecha [r]), Tepiman (O’odham, Tepehuan (Southeastern)), Tequistlatecan (Chontal (Highland), Chontal (Lowland)), Tlingit, Tonkawa (Oxönd), Tohono O’odham (Tolteca), Totonacan (Tepehua (Tlachichilco), Totonac (Misantla)), Tsamosan (Chehalis (Upper)), Tsimshianic (Gitksan, Tsimshian (Coast)), Tunica, Wappo, Washo [r], Wintu (Wintu [n]), Yokuts (Yawelmani [r]), Yuchi, Yukian (Yuki [ra]), Yuman (Diegueño (Mesa Grande) [rd], Hualapai, Maricopa, Paipai [r]), Yurok, Zapotecan (Chatino (Yaitpepec) [ra], Zapotec (Isthmus), Zapotec (Mactuiltianguis), Zapotec (Quiegolani), Zapotec (Tlacolulita), Zapotec (Yalálag), Zapotec (Yatzachi), Zapotec (Zoogocho)), Zuni.

South America: Alacalufan (Qawasqar), Arauan (Jamamadi [r], Paumari [r]), Araucanian (Mapudungun), Arawakan (Apariná, Arawak, Baniwa, Campe (Ajinca), Palikur, Resigaro, Tariana [r], Yavitero), Arikem (Karitiana), Aruak (Ika), Aymaran (Aymara [rad], Jaqaru [r]), Barbacoan (Awa Pit [rad], Tsafiki [ra]), Betoi, Bororo, Cahuapanan (Chayahuita [r], Jebero [r]), Camsá Cariban (Carib, Carib (De’kwana), Hixkaryana, Tiriyo), Chapacura-Wanhan (Wari’), Chimúan (Mochica), Choco (Epena Pedee), Chon Proper (Selknam), Ge-Kaingang (Canela-Krahô, Timbira), Guahiban (Cuiba, Guahibo), Guairuruan (Abipón, Kadiwéu, Toba), Guató, Guaymi (Ngábere [r]), Huitoto (Bora [ra], Huitoto (Murui) [n]), Itonama, Jivaroan (Achuar [n]), Karajá (Javaé [r]), Katukinan (Canamári), Kwazá [ra], Matoacoan (Wichí), Maxakali, Monde (Gavião), Moseténan (Mosetén), Movima, Mundurukú, Mura (Pirahá), Nadahup (Dâw [r], Hup [rad], Nadieb), Nambikuaran (Nambikuára, Sabanê [r]), Páezan (Paez [n]), Panoan (Cashinahua, Shipibo-Konibo, Yaminahua), Paya (Pech), Peba-Yaguan (Yagua), Piaroa, Puelche (Gününa Kiñe), Puinave, Quechuan (Quechua (Huallaga) [n], Quechua (Imbabura) [n]), Rama [ra], Ramarama (Karó (Arára)), Rikbaktsa, Tacanan (Araona, Cavineña), Talamanca (Boruca, Bribri), Trumai, Tucanoan (Barasano
Southeast Asia–Oceania: Aslian (Jahai, Semelai [r], Temiar [rd]), Atayalic (Atayal [r], Seediq), Bahnaric (Chrau, Sapuan, Sre, Stieng), Bodo-Garo (Garo [rd]), Bodic (Belhare [ra], Gurung [ra], Kham [rd], Newari (Kathmandu) [ra], Thulung [ra]), Borneo (Begak-Ida’an, Malagasy [r]), Bugan, Burmese-Lolo (Burmese [rd], Lahu [ra], Lalo [ra], Liangshan Nuosu [r], Lisu [rad], Lolo (Yongren) [r]), Central Malayo-Polynesian (Kambera, Kedang, Ke’o, Ngad’a [r], Tetun), Chamorro, Chinese (Cantonese, Mandarin), Great Andamanese [n], Hmong-Mien (Hmong Njua, Mien), Kadai (Lachi), Kam-Tai (Khamti [ra], Lao, Mulao, Nung (in Vietnam), Phake [ra], Thai, Zhuang (Northern)), Karen (Karen (Sgaw), Kayah Li (Eastern)), Kattic (Pacoh, Sô), Khasian (Khasi [r]), Khmer, Kuki-Chin (Bawm, Lai, Meithei [ra]), Lepcha [rd], Meso-Philippine (Tagalog), Monic (Mon), Nicobarese (Nicobarese (Car)), Northern Philippines (Kalinga (Limos) [n], Kapampangan), Nungish (Dulong [ra]), Oceanic (Ayiwo, Bali-Vitu, Drehu, Fijian, Iduna, Kilivila, Kiribati, Kwaio, Kwamera, Kwara’Ae, Manam, Maori [r], Musom, Oroha, Paamese, Paita, Patep, Rapanui [r], Seimat, Siar, Tawala, Tinrin, Tobati [r], Tolai, Tungak [n]), Paiwanic (Paiwan [r]), Palauan [rd], Palaung-Khuemic (Khmû’, Mlabri (Minor)), Puyuma, Qiangic (Qiang [r]), rGyalrong (Gyarong (Coptse), rGyalrong (Caodeng)), Sama-Bajaw (Bajau), South Halmahera – West New Guinea (Biak, Tabu, Warembori), South Mindanao (Tboli [rad]), Southern Philippines (Tagabawa), Sulawesi (Tukang Besi [n], Wolio), Sundic (Acehnese, Cham (Eastern), Indonesian, Indonesian (Riau), Manadonese [ra], Minangkabau, Nias, Sundanese, Urak Lawoi’), Tani (Bokar [rd], Gallong [rad]), Tsouic (Rukai [rd], Tsou), Tujia, Viet-Muong (Kri, Vietnamese), Yapese.