Identifying semantic role clusters and alignment types via microrole coexpression tendencies

Iren Hartmann[1], Martin Haspelmath[1] and Michael Cysouw[2]
[1]Max Planck Institute for Evolutionary Anthropology / [2]Philipps-Universität Marburg

In this paper, we illustrate a method for identifying clusters of semantic roles by cross-linguistic comparison. On the basis of data from 25 languages drawn from the ValPaL (Valency Patterns Leipzig) database, we show how one can visualize coexpression tendencies using quantitative methods (in particular, multidimensional scaling). Traditionally, the coexpression of semantic microroles (such as the breaker and the broken thing of the 'break' verb, the helper and the helpee of the 'help' verb, etc.) has been studied for particular languages, with generalized macroroles such as "agent", "actor", and "undergoer" being compared across languages in a next step. We set up a conceptual space of 87 microroles based on their coexpression tendencies, i.e. the extent to which they are expressed identically (via flagging and indexing) across our languages. The individual coding means (cases, adpositions, index-sets) can then be mapped onto this conceptual space, revealing broader alignment patterns.

Keywords: microrole, coexpression, semantic map, alignment type, valency

1. Introduction

Semantic role notions are widely used by linguists to capture generalizations of argument coding and argument behavior across verbs within a single language. Thus, in German the distribution of Nominative and Accusative arguments in two-place verbs like *zerbrechen* 'break', *schlagen* 'beat', and *essen* 'eat' is not random, but follows a semantic regularity: The agent (breaker, beater, eater) is coded by the Nominative case, and the patient (broken, beaten, eaten) argument is coded by the Accusative case. However, a precise semantic characterization of roles like "agent" and "patient" is not easy, and in practice linguists have often followed a

Studies in Language 38:3 (2014), **463–484**. DOI 10.1075/sl.38.3.02har
ISSN 0378–4177 / E-ISSN 1569–9978

semasiological (form-to-function) approach, so that they have (often unwillingly) ended up describing patients as whatever Accusative arguments share semantically.

But it turns out that cross-linguistic comparison is difficult when using such semasiologically defined roles. Thus, in German one would exclude the non-agent arguments of *helfen* 'help', *folgen* 'follow', and *dienen* 'serve' from the "patient" role because they are expressed as Dative arguments, while in English the corresponding roles are naturally included. German patients and English patients are therefore no longer directly comparable, and these roles cannot be used readily to compare languages. This situation has given rise to the well-known problems in applying semantic roles (especially role hierarchies) across languages (e.g. Newmeyer 2005: 215–220).

In this paper, we adopt a different approach to cross-linguistic comparison. Instead of starting with language-internal generalizations about roles and then moving to the cross-linguistic level, we show that one can usefully compare languages at the level of the roles of individual verbs (microroles). Semantic role clusters can then be identified by studying cross-linguistic coexpression tendencies, i.e. the ways in which the individual microroles cluster with respect to their coding across a range of diverse languages. Like the more traditional approach sketched in the first paragraph, our method assumes that the coexpression of roles is not random, i.e. that it is not an accident that the breaker, the beater, and the eater are coded alike (= coexpressed) in German and English, and that the helpee, the followee, and the servee are coexpressed in German (by Dative case). Accidental homonymy may exist, of course, but repeated coexpression of the same notional elements across many different languages must indicate similarity of meaning (e.g. Haiman 1974, Wälchli 2009, Wälchli & Cysouw 2012). Our approach is thus a variant of the semantic-map method, applied to microroles of individual verbs and their coding by argument flagging (= case or adpositional marking) or argument indexing (= cross-referencing and "person agreement"). This study does not take word order as a means of coding arguments into consideration, as word order is far more difficult to compare across languages and will have to be left for future research.

2. Three levels of semantic role granularity

There are at least three different ways in which role terminology can be used (and has been used in the literature). These different approaches define roles at three different levels of granularity. First, roles can be formulated at a verb-specific level. For example, verbs like 'hit', 'hug', and 'fear' involve different verb-specific role pairs, namely 'hitter/hittee', 'hugger/huggee', and 'fearer/fearee', respectively.[1] We will refer to such roles as microroles.

Secondly, one can define roles such as 'agent', 'patient', or 'experiencer' at an intermediate level — more abstract than microroles, but still close to the semantic content. For example, the 'agent' might subsume the microroles 'hitter' and 'breaker', the patient might subsume such roles like 'breakee', 'hittee', and 'huggee', and the 'experiencer' might be represented by the 'fearer' in our example. We will refer to the roles at this intermediate level as mesoroles.

Thirdly, roles can be defined at a level that is close to the level of argument realization itself (especially by means of flagging or indexing). From this perspective a role like 'Actor' will subsume the agent and the experiencer from the intermediate level and the 'Undergoer' will subsume the patient and other roles, as for example 'stimulus'.[2] Roles at this higher level of abstraction have been called macroroles (Van Valin 2005) (or hyperroles, Kibrik 1997). The way in which microroles can be collapsed into mesoroles, and these in turn can be collapsed into macroroles is shown nicely by Van Valin (2005: 54); see Figure 1 as an illustration.[3]

Each of these three levels has its problems and limitations. The limitation of the verb-specific level of microroles is that it has no language-internal generality at all. At the other end of the spectrum, the problem with the argument-realization level of macroroles is that it has (too) little cross-linguistic generality. As a result, linguists have typically worked at the level of mesoroles. This level appears to allow for generalizations within a language (allowing one, for example, to account for the similar behavior of 'break', 'hit', and so forth), but it also allows for cross-linguistic comparison. For example, it becomes possible to express many salient differences between languages, such as the contrast between experiencer-subject constructions (e.g. *I like it*) and experiencer-object constructions (e.g. *it pleases me*). Especially for textbook accounts, mesoroles work fairly well, and they have become the best-known type of roles.

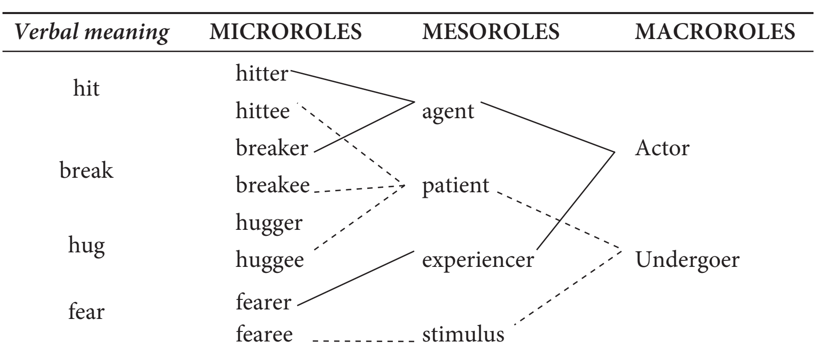


Figure 1. Hierarchy of role concepts (adapted from Van Valin 2005: 54)

However, as we saw in the previous section, mesoroles are still strongly language-specific, and there are many roles that do not fit well into the established categories, e.g. the roles of the boldfaced arguments in (1).

(1) a. *The women searched for* ***the man****.* (searched-for thing)
b. *The women searched* ***the woods*** *for the missing child.* (searched place)
c. *The boy looked at* ***the girl****.* (looked-at entity)
d. ***The old man*** *coughed.* (cougher)
e. ***My arm*** *hurts (me).* (pain locus)
f. ***This man*** *is a hunter.* (hunter)

Verbs such as these may be in the minority, but they show that mesoroles, though commonly used, are not sufficient, whether for cross-linguistic comparison or for language-particular description. Moreover, it appears that those verbs and verb-specific roles that are the most problematic for description are also the ones that differ the most across languages, i.e. while these problems are precisely those cases in which comparison would be the most important, they have so far been intractable.

In this paper, we only use microroles to compare languages.[4] While we are fully aware that our method does not make a particular contribution to the question of how best to describe individual languages, some of the problems in earlier work have stemmed from the misguided attempt to use the same notions for description and for comparison. The microroles that we use here are thus intended exclusively as comparative concepts (Haspelmath 2010). Our approach allows for comparison on the basis of these very basic comparative concepts and does not rely on language-internal generalizations beyond the individual verb and its argument coding.

3. From macrorole alignment to microrole coexpression

The coexpression[5] of different roles by argument-coding elements (case markers, adpositions, person indexes) has been one of the central concerns of typology since the 1970s. This goes by the name of "argument alignment" and has normally worked with prototype-based macroroles such as A, S, P, T, and R (Dixon 1979, Comrie 1981, Haspelmath 2011). The two major types of monotransitive alignment are the accusative type, where intransitive S is coded like A (= coexpressed with A) but differently from P, and the ergative type, where intransitive S is coded like P (= coexpressed with P) but differently from A. This is generally represented in a semantic-map-like diagram as shown in Figure 2a. This analysis works well for prototypically transitive clauses that have an A and a P argument, but the approach

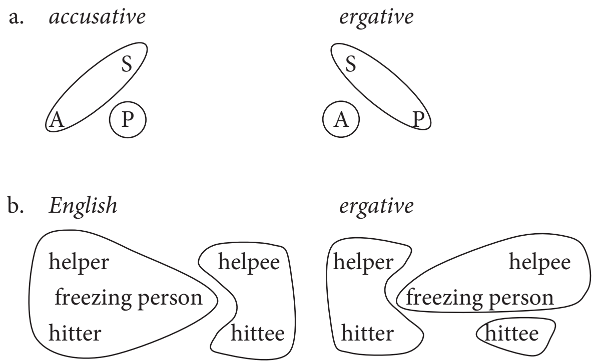


Figure 2. Examples of macrorole (a.) and microrole (b.) argument alignments

has nothing to say with regard to other kinds of two-argument verbs. Likewise, the variation within single-argument verbs (agentive-patientive and similar systems, cf. Donohue & Wichmann 2008) is sometimes accommodated by distinguishing between S_a and S_p, but these two "sub-macroroles" are not particularly homogeneous across languages.

In order to capture a broader range of cross-linguistic differences in argument coding, we need to extend the alignment or coexpression approach to a larger set of roles that can be clearly identified cross-linguistically. This is what we set out to do in the present paper: We study the alignment of the coding of 181 different microroles, which belong to 87 different verb meanings (intended to be representative of the entire verbal lexicon). For example, consider the three aforementioned verb meanings 'help', 'hit', and 'freeze', which behave differently in English and German:

(2) English
   a. *She*$_{\text{NOM}}$ *helps me*$_{\text{ACC}}$.
   b. *She*$_{\text{NOM}}$ *hits me*$_{\text{ACC}}$.
   c. *She*$_{\text{NOM}}$ *is freezing.*

(3) German
   a. *Sie*$_{\text{NOM}}$ *hilft mir*$_{\text{DAT}}$. 'She helps me.'
   b. *Sie*$_{\text{NOM}}$ *schlägt mich*$_{\text{ACC}}$. 'She hits me.'
   c. *Ihr*$_{\text{DAT}}$ *ist kalt.* 'She is freezing.'

The coexpression of the five roles is rather different in English and German, as can be seen in Figure 2b. Some further cross-linguistic differences of this sort can be seen in Table 1, which compares the coding sets[6] of seven microroles in three other languages. We see that only the 'hitter' and the 'helper' are coexpressed in all three languages (see the grey shading). The 'hittee', 'likee', and the 'helpee' are coexpressed in two languages (Hoocąk and Chintang), and the 'liker' and the 'helpee'

Table 1. Seven microroles with associated coding sets in three different languages[7]

microrole	Icelandic	Hoocąk	Chintang
hitter	NP-nom & sbj.V	act.V	NP-erg & V.sbj
hittee	NP-acc	und.V	NP-abs & V.obj
liker	NP-dat	act.V	NP-erg & V.sbj
likee	NP-nom & sbj.V	und.V	NP-abs & V.obj
throw goal	*inn um*+NP-acc	NP+*eeja*	NP-abs & V.obj
helper	NP-nom & sbj.V	act.V	NP-erg & V.sbj
helpee	NP-dat	und.V	NP-abs & V.obj

are only coexpressed in one language (Icelandic). Comparing such pairs of microroles and their coexpression or non-coexpression will be the basis for our study.

It should be noted that the language-specific forms and their labels play no role in our comparisons, only identity of coding between different microroles. If we renamed the Icelandic Nominative and called it "Subjective", and if we did the same with the Chintang Ergative, the two languages would not become more similar. All that counts is the language-internal identity of coding of the microroles.

Our approach thus exemplifies the shift of typology from macro-alignment (using A, S, P etc.) to micro-alignment, i.e. to investigate how individual verbs are treated compared to other individual verbs. This is in line with what Bickel suggests as the characteristic feature of 21st century typology: "Linguistic diversity is captured by large sets of fine-grained variables, not by grand type notions" (Bickel 2007: 245). Our paper thus continues a line of research that was begun by Nichols (2008) for single-argument verbs, but we apply it to a broad cross-section of verbal meanings.

4. Microrole coexpression in 25 languages

Our data consists of coding elements (flags and index-sets) used in a comparable sample of verbs corresponding to 87 verb meanings across 25 genealogically diverse languages from different parts of the world. The data was contributed by language experts participating in the Leipzig Valency Classes Project,[8] and was collected via an electronic database questionnaire, a special database created with FileMaker Pro™. The data is directly accessible in a structured way. It is also comparable, as the verbs were elicited in response to a list of 87 pre-defined verbal meanings, and also because the valency information was recorded in a standardized way (to access the data, please go to www.valpal.info (Hartmann et al. 2013)).

Our 87 verb meanings can be found in Appendix 1 in a manner similar to how they were presented to the data contributors. They are presented in three columns: (i) the meaning label, (ii) the role frame, and (iii) a "typical context".[9] The data was exported from FileMaker Pro™ to be used in the statistical environment R (R Development Core Team 2012).[10]

The 87 different verb meanings include in total 189 different microroles, but because of missing data (i.e. roles for which we do not have information from more than 5 languages) we only use 181 microroles for the current paper. Because the verb meanings correspond closely across languages, the microroles of the verbs can be seen as equivalent across languages and we can thus compare them in terms of their overt coding properties (flagging and indexing). As noted in the preceding section, despite the diversity of the actual coding elements, we can still ask which of the microroles are coded in the same way (i.e. coexpressed) within each single language. Basically, the number of coexpressions between two microroles, averaged over the 25 languages, provides an estimate of the similarity between the microroles.

This approach yields an objective metric on microroles, motivated by the overt form of the argument coding. Finally, we can then produce a visualization of this metric, representing a semantic map of microroles across languages (cf. Cysouw 2010).

In this paper, the metric on microroles was defined as follows. First, when the coding set of a microrole consists of two coding elements (a flag and an index), we separated each coding set into its elements. For example, in Icelandic, the 'hitter' microrole is coded by Nominative case (a flag) and Subject agreement on the verb (an index-set). The similarity between two microroles within a language was then defined as the number of coding elements (not sets) shared by the two, divided by the maximum number of elements in either of the two. Second, for each pair of microroles, the average similarity was taken from all languages for which data was available. Third, we computed the correlation between each of the language-specific metrics and the average metric over all languages (simply using a Pearson correlation between the similarities). The higher this correlation, the more a specific language looks like an 'average' language (i.e. 'average' for this study). To be able to better distinguish diversity, we then computed a weighted average metric by weighting each language using the negative logarithm of the correlation coefficient (this being a very rough approximation of the amount of information of the average inherent to each language separately).

5. Clustering roles

As a visualization of this weighted average metric on the microroles, we used a classic 'metric' multidimensional scaling (MDS). The first two dimensions of the MDS are shown here in Figure 3, which represents a kind of semantic map. Not all 87 microroles are readable because of overlap, but some major role clusters are still discernible. Basically, what happens is that the average coexpression across many languages identifies clusters of microroles which approximately represent the traditional notion of mesoroles (or even macroroles depending on how large one considers a cluster to be) (see Cysouw 2014+ for a detailed discussion). To the left in Figure 3 there is a cluster of agent-like roles, at the bottom there is a cluster of patient-like roles, and at the top right there is a cluster of instruments. Recipients and goals are found in between patients and instruments.

Figure 3 can be regarded as a more detailed version of Figure 2b above, except that the spatial closeness of the microrole labels in Figure 3 was computed on the basis of the coding sets of the 191 microroles, while the arrangement of the five microroles in Figure 2b was done manually. Thus, while the intermediate position of the 'freezing person' in Figure 2b serves to make the figures easy to read, the fact that the 'fear stimulus' occurs right in the middle of Figure 3 reflects the empirical

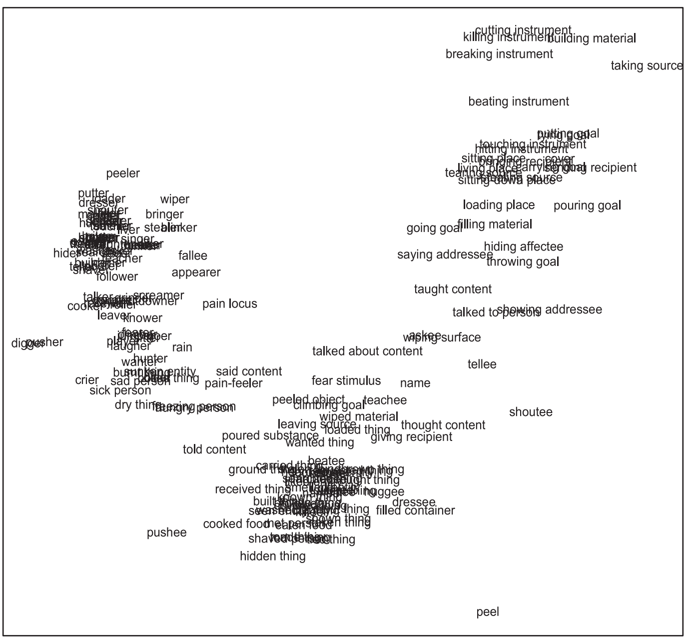


Figure 3. First two dimensions of a multidimensional scaling of the weighted average metric on the 87 microroles, showing clear semantic separation into meso- or macroroles

fact that this role is coded about equally as often (i.e. flagged and/or indexed) as the 'peeler' and 'wiper' (to the upper left) and as the 'tellee' (to the right) and the 'hidden thing' (at the bottom).

To the extent that cross-linguistic coexpression tendencies are due to semantic similarity, we can say that Figure 3 gives us an approximate, but objective representation of the language-independent semantic similarities between the microroles. No subjective judgment of semantic similarity (e.g. between the 'hidden thing' and the 'killee') in any particular language is involved.

6. Mapping languages on the microrole map

In a next step, we can use the two-dimensional layout from Figure 3 as a base map for the comparison of the role marking of individual languages. As a first example of this method, consider the structure of Zenzontepec Chatino role marking in Figure 4. To produce this graphical display, we used the following method. First, the location of the 87 microroles was taken from the multidimensional scaling as shown in Figure 3 (displayed in Figure 4 as small light grey circles). The available data from Zenzontepec Chatino is then plotted on top of this, using different symbols for different language-specific coding elements. In this language, there are only three different argument coding elements: zero (mostly for agent-like arguments) and two flags (prepositions), *jiʔį̀/j-* (for a large variety of semantically disparate arguments, glossed PREP) and *lóʔò* (for instrument-like arguments), as illustrated in (4).

(4) Zenzontepec Chatino (Campbell 2013)
Nkajnyà chu kíʔyùú j-nà niʔií lóʔò lítúu.
CPLV.build NOM.HUM male.DEM PREP-ART house with adobe
'The man built the house with adobe.'

In Figure 4, these three coding elements are identified by squares, circles, and triangles, respectively. The distribution of these three symbols clearly shows clusters, with the squares being used for the agent-like microroles on the left-hand side, the circles for patient-like microroles as well as goal-like, addressee-like, and recipient-like microroles, and the triangle being used for instrument-like microroles in the upper right corner.

To make it easier to see this distribution, we used an R package called Krig to draw lines around the areas of highest density for each coding element. These lines are reminiscent of the boundary lines traditionally drawn in semantic maps (cf. Haspelmath 2003) as well as the contour lines on a traditional geographical map. However, technically speaking, the lines in our figures have a different rationale.

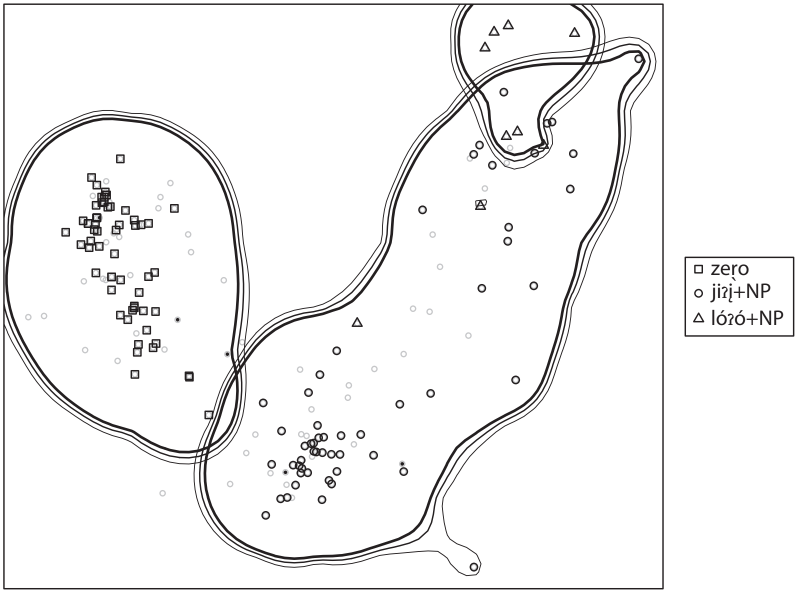


Figure 4. Distribution of the three coding elements in Zenzontepec Chatino

Instead of drawing boundary lines precisely around all points that are coded by a particular coding element, the lines here represent probabilistic indications of the regions in which particular elements predominate. This also explains why there are various coding elements to be found within the 'wrong' lines, for example, some triangles in the area of patient-like coding.

To be precise, the lines represent three different probability distributions (one for each construction) in two-dimensional space, indicating which parts of the figure are more likely to be coded by each element. To show these three probability distributions in one figure we have only drawn lines indicating the probability of 35% (with two thinner lines indicating 32% and 29%, just to visually indicate the gradient nature of these lines). For all lines to be comparable, these probabilities are kept constant throughout all figures in this paper. To infer the probabilities we made use of kriging, a geostatistical method to interpolate distributions in space.[11] In our case, we interpreted the points of Figure 3 as points in space. Then, each point was given a height of one when a specific coding element was present, and a height of zero when a different coding element was attested. Missing data for individual roles was ignored (this can be seen in the grey circles of the base map that are not accompanied by a black circle, triangle, or square). This distribution of high (one) and low (zero) points was then interpolated as 'hills' in space, and the lines were drawn at a height of 0.35, 0.32, and 0.29. In Appendix 2 all of the different distributions of the coding devices from all 25 languages are shown. Because

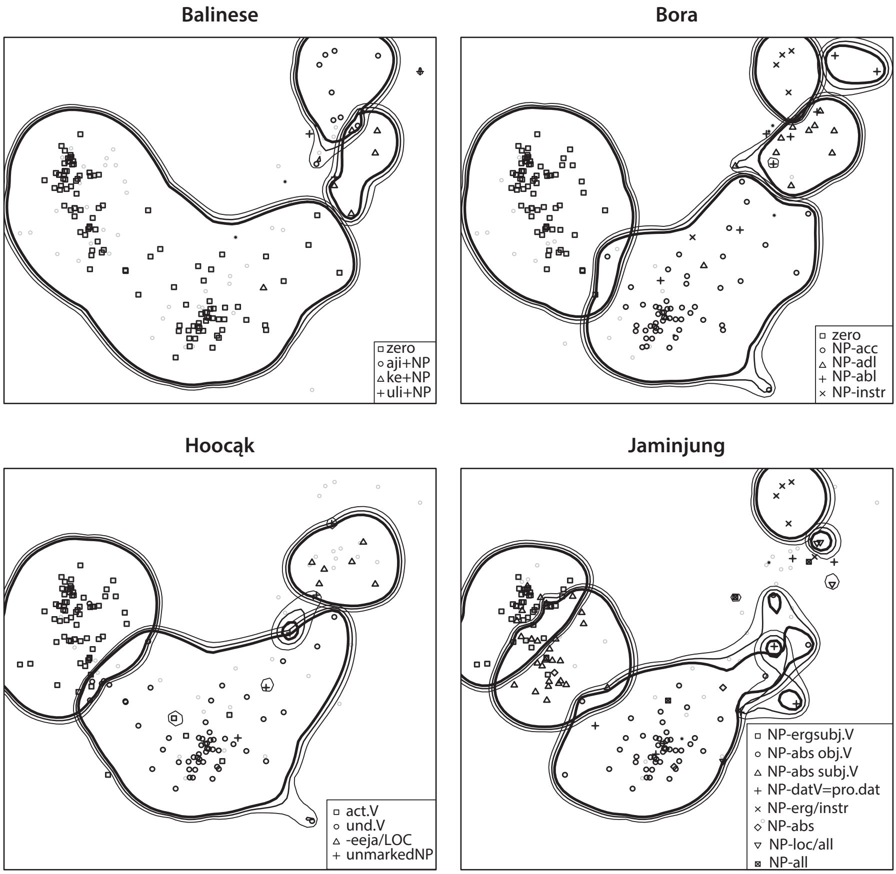


Figure 5. Four additional languages showing different distributional ranges of coding sets

the plots were drawn completely automatically, using exactly the same settings for each plot, the plots are directly comparable across languages.

In Figure 5, we see the distributions of the coding sets in four additional languages, with distribution lines to help us recognize the clusters of coding sets. We see that Balinese is a language that lacks a difference between overt agent and patient coding, while Bora makes a clear agent-patient distinction (Nominative vs Accusative case marking).[12] In Hoocąk, the patient domain is bigger, reflecting its active-stative (Split-S) alignment. Finally, Jaminjung is a double-marking language with a tripartite pattern, distinguishing between transitive agents (Ergative case and Subject indexing), intransitive subjects (Absolutive case and Subject indexing), and transitive patients (Absolutive case and Object indexing). (This kind of tripartite alignment is not generally recognized in the typological literature, because flagging and indexing are considered separately. However, in our approach the full coding set of each microrole is taken into account, both the flag and the index.)

7. Clustering languages: Alignment types

The visualizations shown in Figures 4 and 5 not only present a method for the visual inspection of the structure of different languages, we can also use it to compare languages quantitatively. Instead of attempting to classify the 25 languages into discrete types, we quantify the pairwise similarity between all languages. In this way we are able to investigate global similarities between languages (i.e. to develop a language typology), without ignoring the interesting and important differences between the languages. Instead of discretely classifying languages into types (i.e. making a 'partition'), we provide a metric for languages, specifying precisely how similar each language is to each and every other language. We can then use the metric to provide a partition, or any other simplification of the actual similarities (e.g. a hierarchical clustering as shown in Figure 6 below).

The actual comparison between two languages was performed as follows. Basically, we made language-specific metrics on all 87 microroles, using just the coding of each single language. For each pair of microroles, only the coding similarity from a single language was used. Therefore, when two microroles used the same coding, their difference was zero, while in the case of two completely different coding devices, the difference was set at one. This was performed for all 87*86/2=3741 pairs of microroles, and for all 25 languages. We then used a weighted Pearson correlation to establish the similarity between two language-specific lists of these 3741 pairs of microroles. We used the weighting to control for the fact that the microroles are not equally distributed across all possible functions. As can immediately be discerned from Figure 3, there are many more agent-like and patient-like microroles than instrument-like or goal-like roles. Any unweighted comparison would thus strongly favor whatever differences exist in these highly frequent kinds of roles. To counteract this overrepresentation, we weighted each pair by the inverse of the distance between the roles (i.e. the distance as visualized in Figure 3). This means that pairs of microroles that are far apart are given greater weight in the language comparison, and pairs of microroles that are close together are given less weight.

The resulting metric on the 25 languages is visualized in the form of a hierarchical clustering in Figure 6.[13] A few clusters of languages are indicated in this figure, referring to Appendix 2, where the distributions of the coding devices of all theindividual languages are provided. We essentially find a major division between languages that do not show any coding differentiation between overt agent and patient marking (illustrated above by Balinese), and languages that do have some such division. Within this second group, a large set of languages has an almost identical agent/patient opposition, which primarily represents an accusative alignment pattern, with intransitive subjects classified together with transitive agents

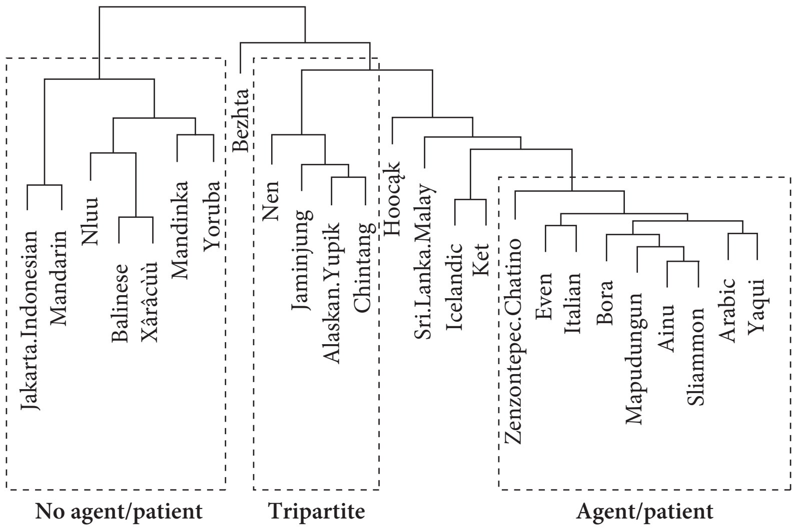


Figure 6. Hierarchical clustering of similarities in microrole coexpression (i.e. alignment similarities)

(illustrated above by Bora). However, the languages between these two groups are the most interesting. Four languages show a tripartite structure, separating different marking for intransitive subjects, transitive agents, and transitive patients (illustrated in Figure 6 by Jaminjung). The other languages show different variants between these clear groups, for example Hoocąk (as seen above) showing a split of the intransitive subjects between agent and patient (compare the division between the agent and patient groups between Bora and Hoocąk to see this difference).

8. Conclusion

Traditionally, the comparison of languages is said to presuppose a thorough language-particular analysis (e.g. Newmeyer 1998: §6.5). But language-particular analysis involving abstract notions such as mesoroles and macroroles (or even cross-linguistic grammatical relations such as subject and object) is highly problematic, as these abstract analytic notions are not really comparable across languages.

In this paper, we have shown that it is possible to arrive at a role clustering and an alignment typology without presupposing any language-particular analysis beyond the level of the individual verb, and without any subjective judgment of semantic similarity. We only made use of information on argument coding (flagging and indexing) for each microrole of 87 verbs in 25 languages from around the world, allowing us to measure the similarities between the microroles and thus arrive at mesorole-like clusters (Figure 3), and in a next step to show the approximate distribution of coding sets over the microroles. This yields an objective, quantitative alignment typology that is not based on any deeper language-particular analysis.

Notes

1. The role names — like 'hugger' and 'huggee' — make liberal use of English word-formation rules and thus will not necessarily sound idiomatic. We use them to avoid clumsy expressions like 'the participant that hugs', 'the participant that is hugged'.

2. Lehmann (2006) has introduced a third macrorole, the INDIRECTUS, but this has not yet become widely known.

3. Cysouw (2014+) introduces yet another level of granularity of roles, viz. CONTEXTUAL ROLE, which can be added even further to the left in the hierarchy or role concepts from Figure 1. The motivation for this is that even verb meanings are sometimes difficult to compare across languages, and it might be favorable in some situations to look at the coding of a role in a specific context, independently of the actual verb that is being used.

4. One reviewer asked whether our approach presupposes that argument coding is only sensitive to the role meaning of the arguments. The answer is that our comparison only captures meaning-based similarities. Argument coding may be sensitive to other factors, such as the grammatical composition of a verb (thus, causative verbs are more likely to have causee and patient coded alike than monomorphemic verbs, cf. Malchukov 2013). Such similarities are not taken into account here, and are a possible confounding factor, though of course the great majority of our verbs are monomorphemic. Argument coding is indeed also often influenced by the referential properties of the argument (e.g. definiteness), but such properties are ignored in our coding sets (cf. Haspelmath & Hartmann 2014+).

5. The term *coexpression* is used here for the relation between a linguistic form in a given language and several functions that are expressed by different forms in some other language. It seems that this is a terminological innovation of the present paper, but we regard it as a very natural one. It was inspired by François's (2008) term *colexicalization* (for lexical coexpression). The term is convenient in a cross-linguistic context because, unlike terms such as multifunctionality or polysemy, it is completely neutral as to the implications for the language-specific description.

6. By CODING SET we refer to the sum of CODING ELEMENT(S) that mark an argument, i.e. a flag (case/adposition) on the NP and/or an index-set (such as subject agreement, object indexing, etc.) on the verb.

7. The abbreviations used in Table 1 stand for the following: abs = Abolutive; acc = Accusative; act = Actor indexing; dat = Dative; erg = Ergative; obj = Object indexing; sbj = subject indexing; und = undergoer indexing. Words in italics are language-specific adpositions.

8. The languages investigated and their respective data contributors are the following: Ainu (Anna Bugaeva), Balinese (Masayoshi Shibatani & Ketut Artawa), Central Alaskan Yupik (Osahito Miyaoka), Modern Standard Arabic (Csilla Kász), Bezhta (Zaira Khalilova & Bernard Comrie), Bora (Frank Seifart), Chintang (Robert Schikowski & Balthasar Bickel), Even (Andrej Malchukov), Hoocąk (Iren Hartmann), Icelandic (Jóhanna Barðdal), Italian (Michela Cennamo), Jakarta Indonesian (Thomas Conners & David Gil), Jaminjung (Eva Schultze-Berndt), Ket (Elena Krjukova & Edward Vajda), Mandarin (Guohua Zhang & Bingfu Lu), Mandinka (Denis Creissels), Mapudungun (Fernando Zúñiga), Nen (Nicholas Evans), N|uu (Martina Ernszt, Alena Witzlack-Makarevich & Tom Güldemann), Sliammon (Honore Watanabe), Sri Lanka Malay (Sebastian Nordhoff), Xârâcùù (Claire Moyse-Faurie), Yaqui (Zarina Estrada, Mercedes Tubino, and Jesús Francisco Villalpando), Yoruba (Joseph Atoyebi), Zenzontepec Chatino (Eric Campbell). We are very much indebted to all our data contributors. Without their expertise and data this study would not have been possible.

9. Since the English verbs used as meaning labels sometimes have different meanings, we have added a sentence for each verbal meaning that makes the intended meaning clear. These sentences are not crucial, they are just intended to help the contributors find a context for their counterpart.

10. We would like to express our sincere gratitude to Hans-Jörg Bibiko for all of his help in this area.

11. We used the *Krig* function for the library *fields* (Furrer et al. 2012) as provided for the statistical environment R (R Development Core Team 2012).

12. Of course, this does not necessarily mean that agents and patients are more difficult to distinguish by Balinese hearers, because word order is an additional coding device that we did not take into account. Thus, we are not necessarily claiming that agents and patients are semantically more similar in Balinese than in Bora. All we are claiming is that they are treated more similarly in terms of flagging and indexing.

13. For this hierarchical clustering we used the function *hclust* from the base package of the statistical environment R. Specifically, we used the *complete* method for the clustering here.

References

Bickel, Balthasar. 2007. Typology in the 21st century: Major current developments. *Linguistic Typology* 11. 239–251. DOI: 10.1515/LINGTY.2007.018

Campbell, Eric. 2013. Zenzontepec Chatino valency patterns. In Iren Hartmann, Martin Haspelmath & Bradley Taylor (eds.), *Valency Patterns Leipzig*. Leipzig: Max Planck Institute for Evolutionary Anthropology. (http://valpal.info/languages/zenzontepec-chatino)

Comrie, Bernard. 1981. *Language universals and linguistic typology: Syntax and morphology*. Oxford: Blackwell.

Cysouw, Michael. 2010. Semantic maps as metrics on meaning. *Linguistic Discovery* 8(1). 70–95.

Cysouw, Michael. 2014+. Inducing semantic roles. In: Silvia Luraghi (ed.), *Semantic roles and semantic maps*. Amsterdam: Benjamins.

Dixon, R.M.W. 1979. Ergativity. *Language* 55. 59–138. DOI: 10.2307/412519

Donohue, Mark & Søren Wichmann (eds.). 2008. *The typology of semantic alignment*. Oxford: Oxford University Press. DOI: 10.1093/acprof:oso/9780199238385.001.0001

Furrer, Reinhard, Douglas Nychka & Stephen Sain. 2012. Fields: Tools for spatial data. *R package*.

Haiman, John. 1974. Concessives, conditionals, and verbs of volition. *Foundations of Language* 11(3). 341–359.

Hartmann, Iren, Martin Haspelmath & Bradley Taylor (eds.). 2013. *Valency Patterns Leipzig*. Leipzig: Max Planck Institute for Evolutionary Anthropology. (http://www.valpal.info)

Haspelmath, Martin. 2003. The geometry of grammatical meaning: Semantic maps and cross-linguistic comparison. In Michael Tomasello (ed.), *The new psychology of language: Cognitive and functional approaches to language structure*, vol. 2, 211–242. Mahwah, NJ: Erlbaum.

Haspelmath, Martin. 2010. Comparative concepts and descriptive categories in cross-linguistic studies. *Language* 86(3). 663–687. DOI: 10.1353/lan.2010.0021

Haspelmath, Martin. 2011. On S, A, P, T, and R as comparative concepts for alignment typology. *Lingustic Typology* 15(3). 535–567.

Haspelmath, Martin & Iren Hartmann. 2014+ (to appear). Comparing verbal valency across languages. In Bernard Comrie & Andrej L. Malchukov (eds.), *Valency classes: A comparative handbook*. Berlin: De Gruyter Mouton.

Kibrik, Aleksandr E. 1997. Beyond subject and object: Toward a comprehensive relational typology. *Linguistic Typology* 1(3). 279–346. DOI: 10.1515/lity.1997.1.3.279

Lehmann, Christian. 2006. Participant roles, thematic roles and syntactic relations. In Tasaku Tsunoda & Taro Kageyama (eds.), *Voice and grammatical relations: In honor of Masayoshi Shibatani*, 153–174. Amsterdam: Benjamins. DOI: 10.1075/tsl.65.10leh

Malchukov, Andrej L. 2013. Alignment preferences in basic and derived ditransitives. In Dik Bakker & Martin Haspelmath (eds.), *Languages across boundaries: Studies in memory of Anna Siewierska*, 263–289. Berlin: De Gruyter Mouton.

Newmeyer, Frederick J. 2005. *Possible and probable languages: A generative perspective on linguistic typology*. Oxford: Oxford University Press. DOI: 10.1093/acprof:oso/9780199274338.001.0001

Newmeyer, Frederick. 1998. *Language form and language function*. Cambridge, MA: MIT Press.

Nichols, Johanna. 2008. Why are stative-active languages rare in Eurasia? A typological perspective on split-subject marking. In Mark Donohue & Søren Wichmann (eds.), *The typology of semantic alignment*, 121–140. Oxford: Oxford University Press. DOI: 10.1093/acprof:oso/9780199238385.003.0005

R Development Core Team. 2012. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org/.

Van Valin, Robert D., Jr. 2005. *Exploring the syntax-semantics interface*. Cambridge: Cambridge University Press. DOI: 10.1017/CBO9780511610578

Wälchli, Bernhard. 2009. *Motion events in parallel texts: A study in primary-data typology*. Bern: University of Bern Habilitationsschrift.

Wälchli, Bernhard & Michael Cysouw. 2012. Lexical typology through similarity semantics: Toward a semantic map of motion verbs. *Linguistics* 50(3). 671–710. DOI: 10.1515/ling-2012-0021

Appendix 1: The 87 verb meanings and microroles

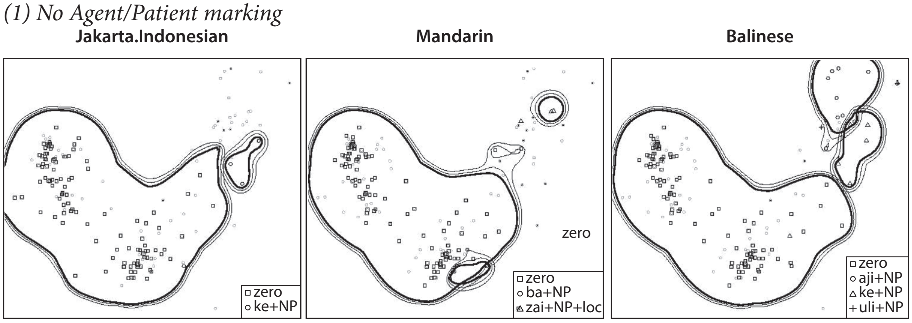
Meaning label	Role frame	Typical context
APPEAR	S appears	A deer appeared (in the road).
ASK FOR	A asks (X) for Y	The boy asked his parents for money.
BE A HUNTER	S is a hunter	This man is a hunter.
BE DRY	S is dry	The ground is dry.
BE HUNGRY	E is hungry	The baby is hungry.
BE ILL	E is sick	The little boy is sick.
BE SAD	E is sad	The little girl was sad.
BEAT	A beats P (with I)	The boy beat the snake with a stick.
BLINK	S blinks	I blinked (my eyes).
BOIL	S boils.	The water is boiling.
BREAK	A breaks P (with I)	The boy broke the window with a stone.
BRING	A brings P T	The girls brought the old lady flowers.
BUILD	A builds P (out of X)	The men built a house out of wood.
BURN	S burns	The house is burning.
CARRY	A carries T (to X)	The men carried the boxes to the market.
CLIMB	A climbs (up L)	The men climbed (up) the tree.
COOK	A cooks P	The woman is cooking soup.
COUGH	S coughs	The old man coughed.
COVER	A covers P (with X)	The woman covered the boy with a blanket.
CRY	S cries	The baby is crying.
CUT	A cuts P (with I)	The woman cut the bread with a sharp knife.
DIE	S dies	The snake died.
DIG	A digs (for) P	The women are digging for potatoes.
DRESS	A dresses P	The mother dressed her daughter
EAT	A eats P	The boy ate the fruit.
FALL	S falls	The boy fell.
FEAR	E fears M	The man feared the bear.
FEEL COLD	S is cold	I'm cold.
FEEL PAIN	E feels pain in M	My arm is hurting. = I'm feeling pain in my arm.
FILL	A fills P (with X)	The girl filled the glass with water.
FOLLOW	A follows X	The boys followed the girls.
FRIGHTEN	A frightens P	The bear frightened the man.
GIVE	A gives T to R	We gave the books to the children.

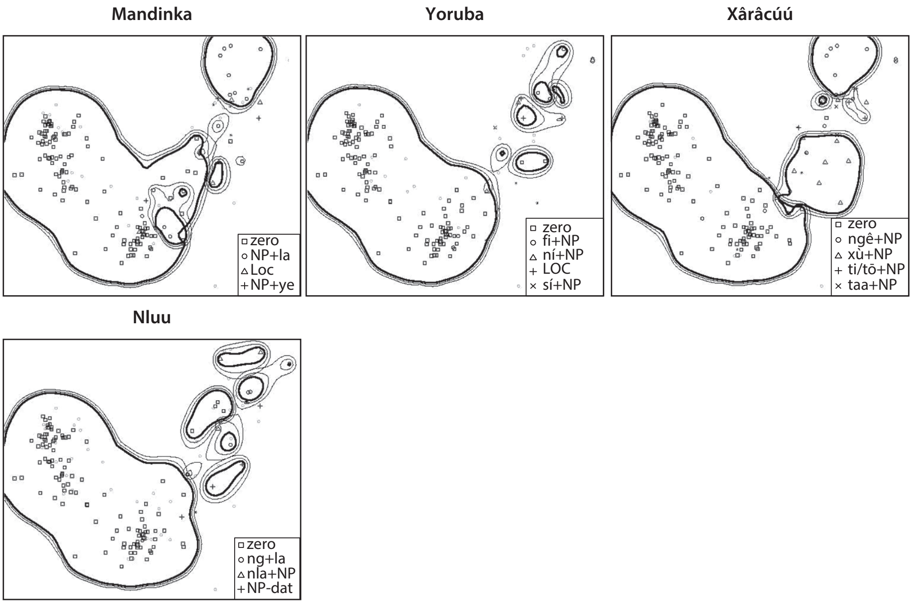
GO	S goes somewhere (L)	The woman went to the market.
GRIND	A grinds P (with I)	The boy is grinding corn with mortar and pestle.
HEAR	E hears M	The boy heard the bear.
HELP	A helps X	I helped the boys.
HIDE	A hides T (from X)	The boy hid the frog from his mother.
HIT	A hits P (with I)	The boy hit the snake with a stick.
HUG	A hugs P	The mother hugged her little boy.
JUMP	A jumps	The girl jumped.
KILL	A kills P (with I)	The man killed his enemy with a club.
KNOW	A knows P	The girl knew the boy.
LAUGH	S laughs	The little girl laughed.
LEAVE	A left L	The boy left the village.
LIKE	E likes M	The boy liked his new toy.
LIVE	S lives somewhere (L)	The old people live in town.
LOAD	A loads T (onto L)	The farmer loaded hay onto the truck. = The farmer loaded the truck with hay.
LOOK AT	A looks at P	The boy looked at the girl.
MAKE	A makes P (out of X)	The men made a house out of wood.
MEET	A meets X	The men met the boys.
NAME	A name X (a) Y	The parents called the baby Anna.
PEEL	A peels (X off) P	The boy peeled the bark off the stick.
PLAY	S plays	The child is playing.
POUR	A pours T somewhere (L)	The man poured water into the glass.
PUSH	A pushes P (somewhere)	The boy pushed the girl (into the water).
PUT	A puts T somewhere (L)	I put the cup onto the table.
RAIN	(it) rains	It rained yesterday.
RECEIVE=GET	R gets T (from X)	The woman got a letter from an admirer.
ROLL	A rolls	The ball is rolling.
RUN	A runs	The horse is running.
SAY	A says "…"(to X)	They said "no" to me.
SCREAM	S screams	The man screamed.
SEARCH FOR	A searches for X	The men searched for the women.
SEE	E sees M	The man saw the bear.
SEND	A sends T (to X)	The girl sent flowers to her grandmother.
SHAVE	A shaves (his beard/hair)	The man shaved his beard/cut his hair
SHOUT AT	A shouts at X	The woman shouted at the children.
SHOW	A shows T (to R)	The girls showed pictures to the teacher.

SING	S sings	The boy sang (a song).
SINK	S sinks	The boat sank.
SIT	S sits somewhere (L)	The children sat on the floor.
SIT DOWN	S sits down (somewhere (L))	The children sat down on the bench.
SMELL	E smells M	The bear smelled the boy.
STEAL	A steals P (from X)	The boys stole apples from their neighbor.
TAKE	A takes P (from X)	The man took the money from his friend.
TALK	A talks (to X) (about Y)	The girl talked to the boy about her dog.
TEACH	A teaches R T	The teacher taught the children English.
TEAR	A tears P (from X)	The girl tore the page from the book.
TELL	A tells (X) Y	The girl told the boy a funny story.
THINK	A thinks about X	The girl thought about her grandmother yesterday.
THROW	A throws T somewhere (L)	The boy threw the ball into the window.
TIE	A ties P (to L) (with I)	The man tied the horse with a rope to the tree.
TOUCH	A touches P (with I)	The boy touched the snake with a stick.
WANT	A wants X	The bear wanted this fish.
WASH	A washes P	The mother washed the baby.
WIPE	A wipes T (off L)	The girl wiped crumbs off the table.

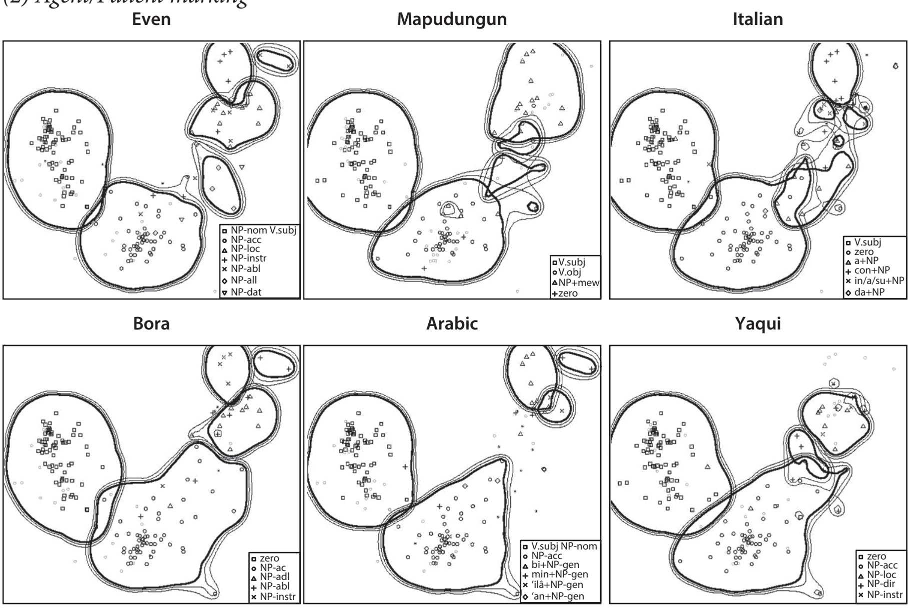
Appendix 2. Distribution of all language-specific coding devices, presented in 5 groupings

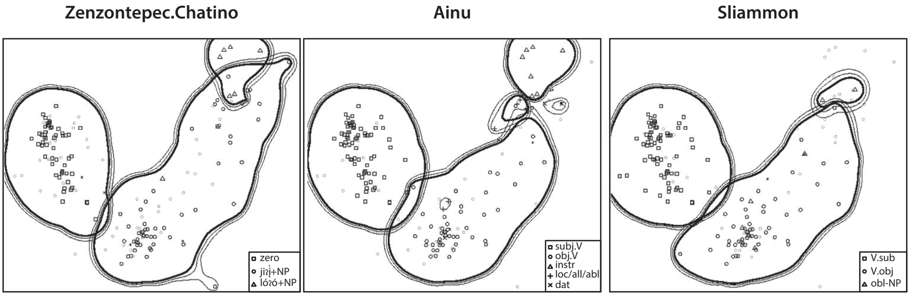
(1) No Agent/Patient marking



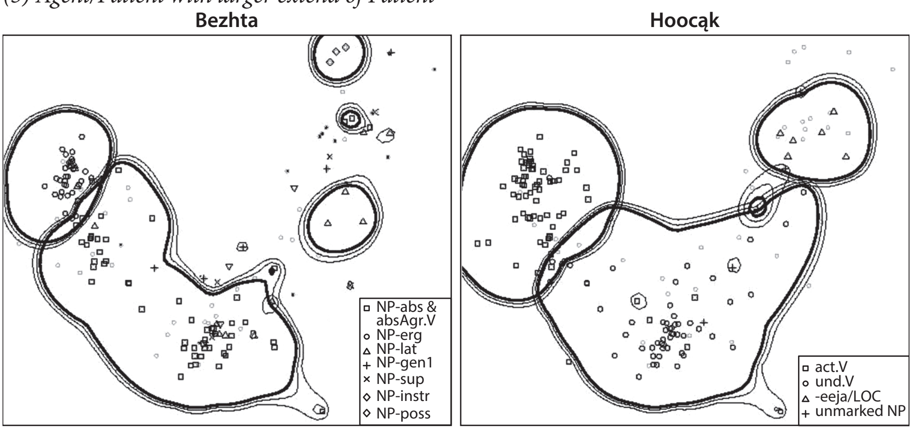


(2) Agent/Patient marking

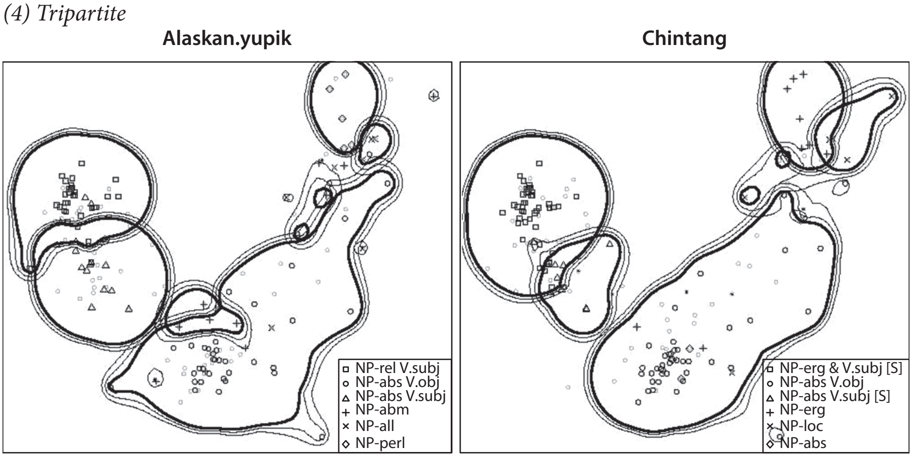




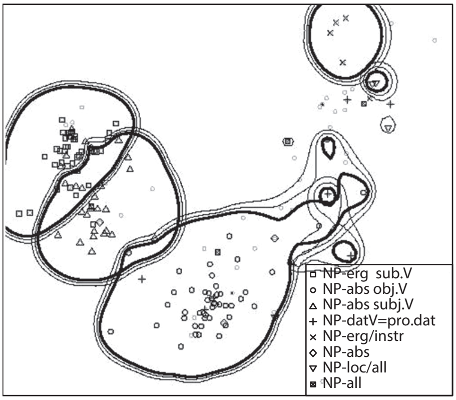
(3) Agent/Patient with larger extend of Patient



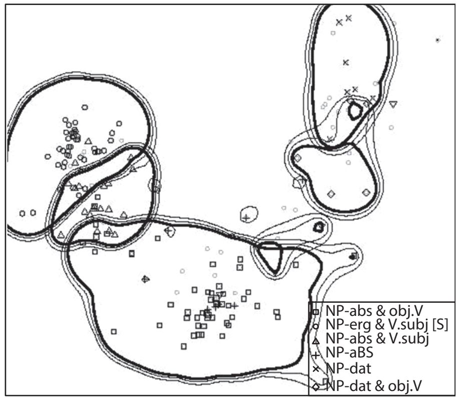
(4) Tripartite



Jaminjung

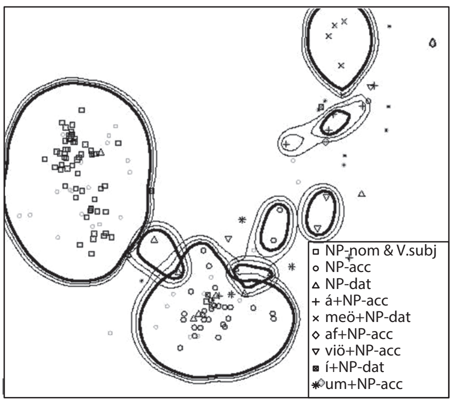


Nen

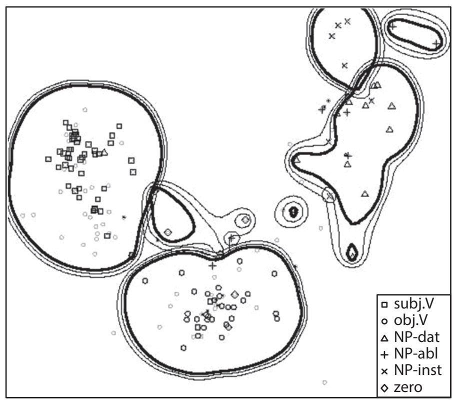


(5) Slightly tripartite

Icelandic



Ket



Sri.Lanka.Malay