Multivariate typology and field linguistics: a case study on detransitivization in Kiranti (Sino-Tibetan)

Bickel, Balthasar

Posted at the Zurich Open Repository and Archive, University of Zurich
ZORA URL: https://doi.org/10.5167/uzh-57331
Accepted Version

Originally published at:
Multivariate typology and field linguistics: a case study on detransitivization in Kiranti (Sino-Tibetan)

BALTHASAR BICKEL

University of Zürich

1 INTRODUCTION

Field linguistics has an intrinsic affinity to the idea of analyzing languages “in their own terms”, and this has traditionally clashed with analytical metalanguages (“theories”) that assume universal, cross-linguistically applicable categories. While in the second half of the last century, field linguists have increasingly adopted universalist metalanguages, more recently, there has been a move away from this again — at least in theory (cf. Haspelmath 2010 for a recent statement), though perhaps less in practice. The current dissatisfaction with universalist metalanguages comes from the observation that their categories often hide substantial cross-linguistic diversity and do not capture well the particularities of individual languages. However, without universally applicable categories, field linguistics becomes unable to contribute to generalizations about language, and, conversely, this insulates linguistic theory and typology from the empirical challenges posed by individual languages. The basic problem therefore is how we can reconcile the need for universally applicable categories with the challenge of particular systems.

In this paper I illustrate the problem with a construction that is frequently found in the Kiranti subgroup of Sino-Tibetan and that seems to escape received notions of traditional, universal metalanguages. I will argue that the construction is sufficiently different from language to language to represent the same cross-linguistic category. However, rather than giving up on a universalist metalanguage, I will argue for an alternative that combines full coverage of language particulars with the goal of developing a universal metalanguage that allows empirically responsible comparisons.

* The research reported here was made possible by Grants No. II/79092 and II/83393 from the VW Foundation and Grant No. BI 799/5-1 from the Deutsche Forschungsgemeinschaft. I would like to thank Martin Haspelmath and Nathan Hill for discussing with me some of the theoretical issues raised here, and Robert Schikowski for checking some of the Chintang data and for comments on an earlier draft.
2 DETRANSITIVIZATION IN KIRANTI

The Kiranti family is a subgroup of Sino-Tibetan and includes about 30 languages spoken in Eastern Nepal. A key characteristic of many languages in this family is that their two- and three-argument predicates can usually alternate between a transitive and an intransitive frame.\(^1\) Consider the following example:\(^2\)

(1) Belhare (ISO639.3:byw; Eastern Kiranti; Bickel 2006)
   a. \(\text{ina-ŋa wa khuiʔ-t-u.}\)
      DEM-ERG chicken[-NOM] [3sA-]steal-NPST-3sO
      ‘That one steals / will steal the chicken.’
   b. \(\text{ina wa khuiʔ-yu.}\)
      DEM[-NOM] chicken[-NOM] [3sS]-steal-NPST
      ‘That one steals chicken.’ (‘S/he is a chicken-stealer’)

In (1a), the verb *khutma* ‘to steal’ is inflected transitively and the A argument (*inaŋa* ‘that one’) appears in the ergative case. In (1b), which I label here a ‘detransitivized’ construction, the same verb is inflected intransitively. The A argument is treated like an intransitive S argument and is therefore assigned nominative case. The semantic difference is that the transitive version implies a specific and quantifiable P referent while the intransitive version suggests a generic or non-quantifiable P referent. Instead of the P argument, also T (theme) or G (goal) arguments of three-place predicates can be the target of detransitivization, depending on the verb class. I will refer to the targeted arguments collectively as ‘objects’.\(^3\)

The basic structure is the same in other languages of the family, but there are important differences. For example, Belhare does not allow modification of the object in the detransitivized construction.

(2) Belhare (Bickel 2006)
   a. \([\text{NP } \text{khaĩ=kha cece}] \ n-cai-t-u.\)
      \[3sS]-good[-NPST]=NMLZ meat[-NOM] 3nsA-eat-NPST-3sO
      ‘They eat (the) good meat.’
   b. *\([\text{NP } \text{khaĩ=kha cece}] \ n-ca-yu.\)
      \[3sS]-good[-NPST]=NMLZ meat[-NOM] 3nsS-eat-NPST
      \(\text{Intended:} \) ‘They eat good meat.’

---

\(^1\) I have not done a complete survey but limit myself to the languages of the Kiranti family that I am most familiar with.

\(^2\) Interlinear glossing follows the Leipzig Glossing Rules, with the addition of the abbreviations \(d\) ‘dual’, \(i\) ‘inclusive’, \(ns\) ‘nonsingular’, and \(s\) ‘singular.

\(^3\) I use S, A, P, T, G as symbols for generalized semantic roles in the sense of Bickel (2011). Unlike in Dixon’s (2010) approach, A and P are semantic roles and are maintained under (anti)passivization. S is the sole argument of a one-place predicate.
c. \([ N \, cece] \) \( n-ca-yu. \)
   meat[-NOM] 3nsS-eat-NPST
   ‘They eat meat.’

Thus, a detransitivized object lacks the phrase-structural properties of a full-fledged NP. The same constraint holds for Limbu (ISO639.3:1if; Angdembe 1998), a language from the same Eastern subgroup as Belhare.

However, another language from the same subgroup, Chintang does allow modification of detransitivized objects:

(3) Chintang (ISO639.3:ctn; Eastern Kiranti; elicited data, 2008)
   a. \( huĩsa-ŋa \) \( [NP \, the=go \, khim] \) \( copt-o-ko. \)
      DEM-ERG big=NMLZ person[-NOM] [3sA-]look-3sO-NPST
      ‘He looks at the big house.’
   b. \( hungo \) \( [NP \, the=go \, khim] \) \( cop-no. \)
      DEM[NOM] big=NMLZ house[-NOM] [3sS-]see-NPST
      ‘He looks at big houses.’

In both Belhare and Chintang it is possible to drop the detransitivized objects, given a suitable context. The following data from Schikowski (in prep.) provide a minimal pair from our corpus (Bickel et al. 2011):

(4) Chintang
   a. \( theke \) \( kham \) \( a-tad-o-ko \) \( utti? \) \[CLLDCh2R06s02.1402\]
      why soil[NOM] 2sA-bring-3sO-NPST that.much
      ‘Why do you bring that much soil?’
   b. \( hunci-jhani=yaŋ \) \( taʔ-no=naŋ. \) \[CTN_Fut_Pln.493\]
      3dPOSS-family[-NOM]=ADD [3sS-]bring-NPST=but
      ‘But their family also brings (some dowry into marriage)’

(4a) represents a fully transitive sentence, with a quantified object (‘that much soil’, represented by a discontinuous constituent \( [kham... \) \( utti] \)). In (4b), the construction is detransitivized, and this implies a non-quantifiable object. This argument is not overt, but the wider context of the utterance makes it clear that it must be the dowry that a women’s family is expected to pay.

In other Kiranti languages, detransitivized objects are obligatory, however. This is the case in Puma, where we have not found examples with a dropped argument in our corpus (Bickel et al. 2009):

(5) Puma (ISO639.3:pum; Central Kiranti; Bickel et al. 2007)
   a. \( ɲa \) \( redio \) \( en-ɲa. \)
      1s radio hear-1sS.NPST
      ‘I do radio-hearing.’
   b. *\( ɲa \) \( en-ɲa. \)
      1s hear-1sS.NPST
      \( Intended: \) ‘I hear something.’

3
3 POSSIBLE ANALYSES

In the preceding section I labeled the construction ‘detransitivization’, but that is clearly too wide a term. Can we analyze the construction as an instance of a more specific category? Various possibilities offer themselves (and have indeed been proposed by various authors).

A first option is to analyze detransitivization as an instance of **optional (differential) object agreement**, similar to what is known for example from Bantu languages: in one variant, the verb agrees with the object, in the other it doesn’t. However, this misses the effect on case assignment: under detransitivization, the A argument is not assigned the same case (ergative) as in the intransitive version, but it is consistently assigned the nominative case. This is unexpected if the difference between the constructions lies only in the agreement pattern. One could of course claim that nominative case assignment is a side-effect of intransitive agreement or indeed licensed by intransitive agreement (which always assigns nominative case in the Kiranti languages surveyed here). However, an analysis in terms of optional object agreement would leave unexplained why the object not only loses its ability to trigger agreement, but that in some languages it also becomes obligatorily overt (Puma) or that in other languages it no longer constitutes a full-fledged NP (Belhare, Limbu). The changes in object properties suggest that detransitivization involves a more fundamental construction than just optional agreement.

One hypothesis is that the construction involves **antipassivization** (Weidert & Subba 1985:106). Immediate support for this comes from the fact that the A argument is treated like an S argument in terms of case and agreement, i.e. it appears to become more similar to what one might think of as a canonical ‘subject’. However, this analysis is in conflict with the properties of the object. Under an antipassive analysis, one would expect the object to lose not only some but all object properties, or at least the most important ones. However, while some properties indeed get lost, other key properties are retained. Most importantly, object case-marking stays the same under detransitivization: the detransitivized object is assigned the same nominative case that objects are also assigned in fully transitive clauses. The only exception to this is Puma, where detransitivization constrains the case-marking possibilities of the object: while in transitive clauses, the object appears in the dative under specific conditions, detransitivization does not allow datives (Bickel et al. 2007). Crucially, however, the constraint does not lead to the assignment of, say, a locative case. This is unlike what one would expect from an antipassive.

Two other important properties are fully retained by detransitivized objects, and they behave in these regards like all other objects of these languages: detransitivized objects can be moved to various positions and, consistent with this, they can be relativized on. The following data show this for Puma; the same is true of all other Kiranti languages where I had the opportunity to check this (Chintang, Belhare):
These data also call into question an alternative analysis of the constructions in terms of compounding (Weidert & Subba 1985:107) or incorporation (Angdembe 1998). Incorporation is also an unlikely option because all inflectional material (prefixes, suffixes) surround the verb stem alone. In none of the languages surveyed here is the detransitivized object inserted inside the inflectional word. This is in contrast with verb compounding, a widespread pattern in Kiranti (e.g. Belhare ṣṇ-ᵽₜᵽₜ-ᵽₜ ‘they fried it and brought it down’, where the stems ṣᵽᵽᵽ ‘fry’ and ᵽₜᵽ ‘bring down’ are flanked by the agreement prefix ṣṇ- ‘3nsA’ and the past tense suffix -ᵽₜ).

Another possibility consists in analyzing detransitivization in terms of a lexical alternation, similar to such alternations as English to eat an apple vs. to eat, i.e. to what is known as ambitransitivity or lability (van Driem 1987). To some extent, there are parallels to canonical ambitransitivity examples, as in the contrast between the transitive example (7a) and the detransitivized example (7b). But the fact that the object argument is often overt under detransitivization (as in 7c), limits the parallelism (Schikowski et al. 2010).

The problem is even more severe in other languages, such as Puma, where objects are always overt (cf. the data in 5). In addition, an analysis in terms of a lexical ambitransitive class is in conflict with the observation that, unlike ambitransitivity in other languages, detransitivization is not limited to a lexically enumerable class. There are only specific exceptions of verbs that do not allow detransitivization, such as small lists of deponent and some experiencer verbs.

The preceding discussion suggests that the Kiranti detransitivization construction is a kind of its own: it does not seem to fit any of the categories that traditional analytical metalanguages define as universally applicable. In response, one could
posit a new category, say, “object detransitivization” and add it to the inventory of universally applicable categories. This new category would be instantiated so far only by Kiranti languages, but it is obviously possible that other languages will have the same category. Good candidates are similar constructions in many other Sino-Tibetan languages (e.g. from the Kuki-Chin family: Bickel 2006) or biabsolutive constructions in Nakh-Daghestanian languages (Forker in press). There are probably quite a few more candidates in other parts of the world.

However, it is difficult to define such a new category because there is substantial variation within Kiranti (and no doubt even more variation if we include similar constructions in other languages). The variation affects key properties of the constructions: in some languages, the detransitivized object is obligatorily present (Puma), in others it can be dropped (Chintang, Belhare); in some languages, the object is a full-fledged NP (Puma, Chintang), in others it is a bare noun that cannot be modified (Limbu, Belhare). The variation cross-cuts languages, and this precludes the possibility of positing two sub-types, e.g. one combining full NP status and the possibility of dropping, and one combining bare nouns with obligatory presence.

One could of course declare one property as necessary and the other as incidental, e.g. define “object detransitivization” by a pattern of intransitive case and agreement. The fact that Belhare and Limbu do not allow objects to be modified would then be an accidental additional property. But unlike such notions as antipassivation, analysis of a construction as “object detransitivization” would not be very interesting because it would not make any predictions about additional properties. In response, one could try and add the reduced NP status as a necessary property, assuming Belhare and Limbu to represent canonical “object detransitivization”. But this would of course be arbitrary: one could just as well take the Puma and Chintang patterns as canonical and assume that normally, object detransitivization assigns objects full NP status. The same issues arises when one tries to take either the obligatory or the optional presence of overt object arguments as a defining property of “object detransitivization”: again, the choice is arbitrary.

In summary, the Kiranti detransitivization constructions cannot be subsumed under any received concept and positing a new universal category forces one to make arbitrary choices when defining the category.

4 PARTICULARS AND UNIVERSALS

To solve the problem that we noted in the preceding section, some typologists advocate a strict separation of the metalanguage of language-specific analysis and the metalanguage of comparative research. Under this approach, each language needs its own metalanguage, while the metalanguage of typology operates with broad terms that deliberately gloss over language-specific details (Dryer 1997, Croft 2001, Lazard 2006, Haspelmath 2010): terms like ‘antipassive’ then receive a universal definition that need not match the criteria one needs for describing the facts of particular languages. For example, antipassivization could be defined as a type of detransitivization that either suppresses objects or realizes them as an oblique complement
(Polinsky 2005). Any additional properties, such as the phrase-structural status of or positional constraints on objects are then irrelevant for identifying the construction in a language. In order to better accommodate the Kiranti data, one might posit a typological category of ‘object downgrading’ including any construction that strips objects of at least one property they normally have in transitive constructions of the language. Any further details would again be irrelevant.

Independently of this, one would posit language-specific categories: Puma Object Detransitivization, Chintang Object Detransitivization, Limbu Object Detransitivization etc., and there would be no expectation that these constructions should instantiate the same category. As Haspelmath (2010) points out, the capitalized initials signal that the labels are proper names: just like we don’t expect London in Canada to be the same as London in the UK, we don’t expect Puma Object Detransitivization to be the same as Chintang Object Detransitivization. Or, perhaps more in line with actual practice in the field, one would expect at best Wittgensteinian family resemblances (i.e. the proper names would be more like family than city names). Either way, language-specific categories of this kind are different from the cross-linguistically recurrent categories that are the basis of typological databases and generalizations.

This approach allows one to go on with typology and language-specific analysis, but it cuts all ties between the two. Yet typological notions are often of critical help in language-specific analyses, and it is one of the great steps of progress that typological and theoretical knowledge increasingly informs descriptive groundwork: for example, without the publication of Comrie (1976), descriptive grammars wouldn’t have analyzed aspect in as detailed a manner as they did (and as witnessed by the hundreds of references to Comrie’s book). Without the publication of Foley & Van Valin (1984), the issue of illocutionary scope properties would have had little chance of ever being addressed in descriptive grammars when analyzing clause linkage. In fact, as many early descriptions in the American structuralist tradition testify, any attempt at describing languages purely ‘in their own terms’ risks missing important analytical questions — in short, no analysis can proceed out of a theoretical vacuum.

Also, separating typological from language-specific analysis moves typology and theory ever further away from the empirical facts of individual languages. Thus, theories could go on assuming that a universal inventory including antipassivization, differential object agreement, ambitransitivity and incorporation is sufficient for analyzing all patterns in which object agreement is optional. The details of the Kiranti facts would simply be irrelevant. This creates the illusion of finding more uniformity in the languages of the world than is empirically warranted, and it robs linguistics of a unique chance to explore, measure and explain one of the most basic facts of our species, its unique cognitive flexibility and adaptivity (cf. Evans & Levinson 2009).
5 MULTIVARIATE TYPOLOGY

An alternative solution to the problem of reconciling language particulars and universals is what I call ‘multivariate typology’ (Bickel 2010). The basic idea is that any similarity between two structures (between or within languages) means that the two structures are identical in some regards and different in other regards. For each relevant ‘regard’, one develops a variable (parameter of comparison, tertium comparisonis) and determines the values (properties) that individual constructions have on these variables in a survey of languages. The variables and their values can be defined in terms of universally applicable notions because they only capture a specific aspect of a construction at a time. This is radically different from traditional universal metalanguages for linguistic analysis. In traditional metalanguages, saying that a particular phenomenon instantiates a given category is expected to capture more than one specific property, e.g. saying that something instantiates antipassivization lets us expect a whole range of properties (with regard to case assignment, agreement structure, syntactic behavior of arguments etc.). Indeed, if a traditional concept were to capture only one specific property, it would be uninteresting (and rejected): it would simply label a phenomenon.

A tentative system of variables for capturing the constructions surveyed above is given in Table 1. The symbols S and A stand for generalized semantic roles: S for the sole argument of a one-place predicate and A for the most agentive argument of a predicate that entails more than one semantic role. The symbol O in the table has a different status: it is used here as an abbreviation for the specific grammatical relation that is targeted by the construction. This opens up another complex system of variables, including the range of semantic roles entailed by predicates, the referential status of the argument participating the grammatical relation, the lexical predicate (valency) class that the argument belongs to, and much else beside. A survey of the space of variation is beyond the scope of the present paper, but an analysis is available for Chintang (Bickel et al. 2010; Bickel 2011) and Witzlack-Makarevich (2011) discuss the relevant variables in theoretical perspective.

Another system of variables that is not included in Table 1 concerns the semantics of the construction. Bickel et al. (2007) suggests that the relevant feature for Puma is that detransitivization cuts all possible entailments to the cardinality of the set of object referents. It remains to be seen how well this carries over to other languages. At any rate, the development of a system of semantic variables requires more detailed analyses of the constructions in more languages.

More generally, a table such as the one presented here, only ever reflects a given state of our progress in understanding a particular phenomenon. Most notions that appear in the variables open up new systems of variables (‘NP’, ‘inflectional word’, ‘case’ etc.). But this is simply an invitation for cumulating ever-more precise analyses. It does not mean that the table misses the nature of the constructions it covers

4 An important terminological consequence of this is that one can speak of ‘diagnostics’ or ‘tests’: whether a specific phenomenon instantiates a given universal category can be ‘diagnosed’ via the range of properties that are expected from the category.
<table>
<thead>
<tr>
<th>Variable</th>
<th>Limbu</th>
<th>Chintang</th>
<th>Belhare</th>
<th>Puma</th>
</tr>
</thead>
<tbody>
<tr>
<td>O-agreement</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>no</td>
</tr>
<tr>
<td>S-case (nominative) on A</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>Regular O-cases on O</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>no</td>
</tr>
<tr>
<td>O is obligatorily overt</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>yes</td>
</tr>
<tr>
<td>O is a full-fledged NP</td>
<td>no</td>
<td>yes</td>
<td>no</td>
<td>yes</td>
</tr>
<tr>
<td>Relativization on O</td>
<td>?</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>O is positionally free</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>O and verb form one inflection word</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>no</td>
</tr>
</tbody>
</table>

Table 1: A multivariate typology of the syntax of some Kiranti detransitivization constructions

(in the way a traditional analysis of the constructions as, say, differential agreement can be said to miss the nature of the constructions).

6 DISCUSSION AND CONCLUSIONS

What is the benefit of a multivariate typology of the kind sketched above? The most obvious benefit is for the analysis of individual languages: having a set of variables at our disposal allows one to ask the relevant questions so that the description will cover not only the most apparent properties (like the case and agreement pattern) but will also give information about the phrase structural status of NPs, relativizability, etc. Another benefit is that unlike the classical approaches discussed in Section 3, multivariate typology does not force one to squeeze a construction into a predefined set and instead allows one to capture all relevant properties. Each language-specific construction consists of the vector of values in the typology, the ‘yes’ and ‘no’ responses in the columns of the table. This vector can be named for practical purposes (e.g. the Puma Detransitivization Construction), but in contrast to the approach proposed by Haspelmath (2010) and others, the construction is still defined in terms of universally applicable notions. Similarities can be stated in exact terms (viz. in terms of ‘yes’ and ‘no’ responses that the constructions have across variables), rather than only via vague appeals to family resemblances.

But at the same time as it facilitates language-specific analysis, multivariate typology offers new perspectives for typological research, probing into universal or areal trends. For example, the fact that the Puma Detransitivization construction requires overt object arguments possibly correlates with the fact that in this language, there is an alternative detransitivization strategy which bans overt object arguments (marked by kha- and discussed in Bickel et al. 2007). For this correlation to be a genuine explanation of the relevant properties it needs to be a universal correlation. As
such, we need a theory that predicts the correlation (e.g. by assuming that obligatory presence of an element entails obligatory absence of the same element in an alternative construction with the same function). But, equally importantly, the correlation needs statistical support from a large sample of languages and language families. This is what multivariate typology offers directly: the system of variables in Table 1 can be filled by detransitivization constructions in many languages (including constructions called antipassivization, optional agreement or whatever). The resulting database then allows statistical evaluation of the hypothesis.

Further, the database can be mined for recurrent patterns: patterns that recur universally or patterns that are specific to a linguistic area. In this way, multivariate typology allows one to search for large-scale generalizations and at the same time to conduct language-specific analysis with full respect for the diversity that we encounter in every fieldwork project. This reconciles the demands of field linguistics and of comparative research and ensures much needed cross-fertilization between these two domains of linguistic inquiry.

REFERENCES


corpus of language acquisition by six children (ca. 500,000 words). DOBES Archive, http://www.mpi.nl/DOBES.


Forker, Diana. in press. The biabsolutive construction in Nakh-Daghestanian. *Folia Linguistica* 46.


