

Current approaches in typology and the language-specific vs. cross-linguistic opposition

Diana Forker (University of Bamberg & JCU)
diana.forker@uni-bamberg.de

There is an obvious difference between describing and/or analyzing an individual language and comparing languages thereby searching for cross-linguistic similarities.

- How does typology deal with this?
- What are recurring problems?

1. Multivariate analysis

- Bickel & Nichols (2002), Bickel (2010), Bickel et al. (2014)

A schematic representation:

1. Identify constructions in languages (a form paired with a range of meanings that can be identified through translations) and build up a database

- example: semantic role types and non-default case assignment (Bickel et al. 2014)
- one-place predicates: ‘the one who feels cold’, ‘the one who has fever’, ‘the one who feels sleepy’, etc.
- two-place predicates: ‘the one who sees’, ‘the one who knows’, ‘the one who finds out’, etc.
- build up a database of non-default cases (or case alternations)

2. Find (if there are any) clusters of constructions that are similar to each other by means of statistical analysis

- examine to what extent predicate-specific semantic roles are grouped together by the same non-default cases (or alternations) in each language
- derive from this a measure of dissimilarity of the roles across languages
- the resulting dissimilarity matrix is then examined for statistical clusters
- > clusters with fuzzy boundaries that are centered around a prototype
- e.g. for one-place predicates the minimal number of clusters that gives good results is 3; for the A-type argument of two-place verbs it is 6

3. Check if these clusters meet our expectations and definitions of existing constructions

- any resulting cluster of predicate-specific roles is potentially indicative of cross-linguistically relevant role complexes
- > statistical analysis of similarities and dissimilarities between clusters that might indicate relations between role clusters (and thus in the end support our assumptions about certain semantic roles)

- S-class: relative clear separation between

- experiencer (e.g. *feel like laughing, feel ashamed, feel fear*)
- undergoer of body processes

- but there is no structure within the experiencer cluster (e.g. no division between emotions and sensations)

- A-class: a few clusters emerge, but they are not very well supported, e.g.

- experiencer (*feel irritated, enjoy, annoy*)
- cognition and *perception* (*find, know, hate*)

- no clustering possible for the P-class

→ only a few clusters support our assumptions about semantic roles, e.g. the distinction between undergoers of body processes and experiencers for one-place arguments

- the A class shows a separation of experiencers vs. cognizers/perceivers, although the signal here is not very strong

- the analysis of the T class of arguments fails to support a clear distinction between instruments and themes

- the study suggests new role clusters that are not traditionally assumed (e.g. communicator within two-place predicates)

4. If this is not the case, then either we can redefine our notions or we introduce new notions (if we can make any sense of the clusters) or we just say that the clusters and the traditional notions have nothing to do with each other

Premise/presupposition: if we want to find these clusters, they must

- exist
- be frequent enough to result in a statistical signal

Assumptions

- some notions are needed as starting points (e.g. *non-default case frame*):

-> sometimes mere frequency can help to apply these notions (e.g. default case frames: cases used with the largest, open classes of predicates)

-> sometimes other means are needed, i.e. a *tertium comparationis* (e.g. Dowty's lexical entailments for distinguishing between the two arguments of two-place predicates)

Advantages

- constantly updating: new constructions can be discovered that do not fit into known schemas
- constructions (descriptive notions) are broken down into their smallest units
- exemplar-based definitions that can make it possible to place one's field language within a typology at an early stage of descriptive work and that in turn can help to identify hypotheses for further research (elicitation)
- strictly inductive: not based on prior models of possible and impossible human languages

Problems

- in the beginning relatively time-consuming because each new construction requires review (and possibly revision) of all previous entries
- noise in the data can be a problem for the statistical analysis (e.g. individual predicates may be forgotten in grammars)

- we cannot directly arrive at our concepts, and maybe the concepts are just an invention that is maybe useless (we see patterns where there are no patterns)
- Is statistics the right methods to detect the denotation of our concepts and investigate the meaning of our constructions?

2. Canonical typology

- e.g. Corbett (2005, 2008), Brown, Chumakina & Corbett (2013), Brown & Chumakina (2013)

A schematic representation:

1. Identify canonical instance(s) of a construction (*canonical ideal*)

- e.g. 'true friendship', 'honest person', 'bird', 'cosubordination', 'argument', etc.
- the canonical instances are idealized; they do not need to correspond to anything found in the real world, nor to something in the speaker's mind or brain
- they are theoretical / philosophical / logical concepts

- example: Corbett (2008) investigates canonical morphosyntactic features and takes *case* as an example

2. Take the canonical instances as the ideal endpoints of a scale and build up the possible logical space for a construction (= *base*: the broad space of a linguistic phenomenon) by figuring out the relevant properties (*criteria*) and their more or less canonical values

- e.g. true friends help each other, a honest person always tells the truth, a bird can fly, in a co-subordinated structure ..., an argument must be a term, semantically and syntactically obligatory, etc. ...
- the base can be the result of a long-standing debate, but it can also be a notional starting point
- the dimensions need to be logically independent of each other

- Corbett (2008) provides ten canonical features and values for *case*, e.g.

- Criterion 1 & 2: Canonical features and their values have a dedicated form and they are uniquely distinguishable across other logically compatible features and their values.
 - case syncretism and portmanteau morphemes expressing case and number together are not canonical
 - e.g. Russian instrumental is more canonical than Russian accusative case
- Criterion 3 & 4: Canonical features and their values are distinguished consistently across relevant word classes and across lexemes within relevant word classes.
 - Russian cases are canonical in this respect (they are expressed on nouns, pronouns, adjectives)
- Criterion 5: The use of canonical morphosyntactic features and their values is obligatory.
 - Russian cases are canonical in this respect (their use is obligatory)

3. Place language-specific construction within this space more or less close to the canonical instance(s)

- we take our 'real life' instances and check which properties they have and to which extend

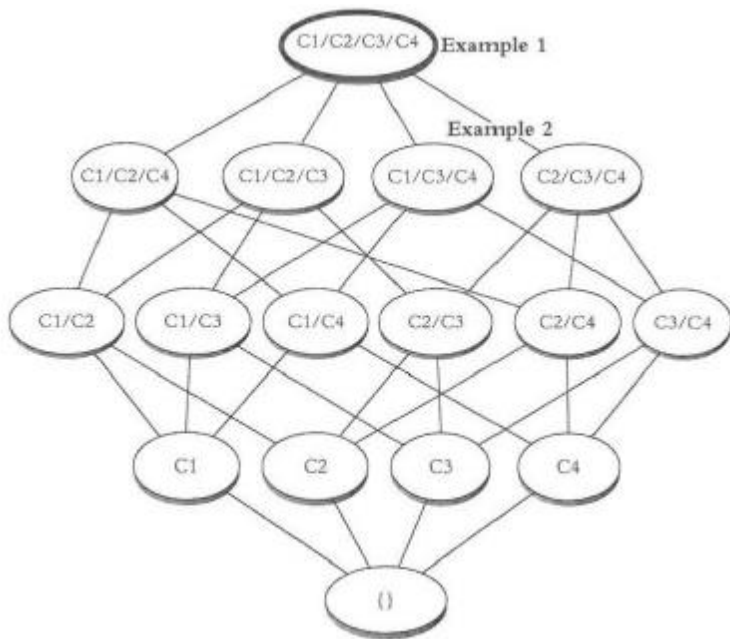
- e.g. my friend Peter helps me sometimes, but not always; a dove can fly but a penguin cannot

- there is only one way of being canonical, but several dimensions of non-canonicity

- In Russian the six traditionally assumed cases are relatively canonical (the instrumental being more canonical than the accusative)

- other 'cases' are less canonical and thus often not assumed to be independent cases (e.g. vocative, second genitive)

Figure 1: Proximity to the canonical ideal (Brown & Chumakina 2013b: 7)



- Premise/presupposition:

- it is possible to define a construction consistently through a multidimensional space of criteria
- no unitary core of meaning is presupposed
- canonical instances can be infrequent, even non-existent
- canonical instances might not be the best-known or clearest examples
- criteria should be such that they can possibly be reused for other investigating other constructions

Other problems

- How do we identify canonical instances and the criteria a priori?

3. Interim summary

- the list of variables and values / criteria is the crucial point

(I) How do we make up the list? Where do the variables/criteria come from?

- bottom-up (multivariate analysis): from data to criteria/variables and values
- top-down, i.e. philosophical/logical (Canonical Typology): from canonical instances and criteria/variables and values to actual instances (data)

(II) What needs to be included and what not? How do we know what the relevant / defining properties are?

- the list and the variables and values /criteria need to be
 - independent
 - complete, but not superfluous (necessary and sufficient)
 - not too short and not too long

(III) What about the distinction between language-specific variables / criteria / categories and cross-linguistic variables/criteria/categories?

4. The language-specific vs. cross-linguistic opposition -- fieldwork and typology

Haspelmath (2010)

- we need to distinguish between two independent levels of linguistic analysis
 - language-specific descriptive categories
 - comparative concepts
- there is a third category, cross-linguistic categories (e.g. Universal Grammar), but they are not suitable for cross-linguistic comparison

Language-specific descriptive categories

- each language has its own categories and can be described in its own terms
- the description depends on the language as much as on the linguist

Comparative concepts

- cannot be right or wrong, but only more or less well-suited to the task (linguist-specific)
- the choice of comparative concepts depends on one's goals, and comparative linguists will always have a multiplicity of goals
- there will never be a single ('standard') list of such concepts (opportunism)
- universally applicable because created by linguists for the purpose of comparison and cross-linguistic generalizations (language universals)
- defined on the basis of

- universal conceptual-semantic concepts
- general formal concepts
- and other comparative concepts

- examples of

- comparative concepts: adjective, verb, lexeme, syntactic pattern, question word, question-word phrase, *clause*, relative clause, reflexive pronoun
- conceptual-semantic concepts: future time reference, property, narrow the reference, *semantic role*, anaphoric
- general-formal concept: (special) position, *clause*, precede, overt

- example of the definition of *adjective*: a lexeme that denotes a descriptive property and that can be used to narrow down the reference of a noun

- no specific terminological convention is proposed

- might in general be good to adopt more semantic terms but this is not a general solution, because comparative concepts are by no means always or even typically purely semantic

- comparative concepts based on fewer factors seem to have greater chances of leading to deeper insights

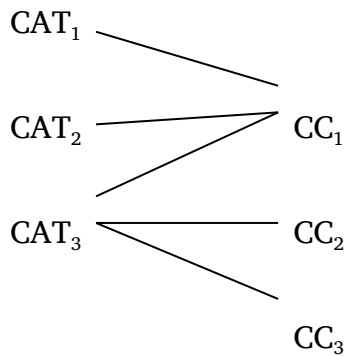
- some terms have a double nature:

- *clause* is a comparative concept and also a general-formal concept
- *semantic role* is a conceptual-semantic concept (and probably also a comparative concept?)

- there is no taxonomic relationship between language-specific categories and comparative concepts, e.g. the Eskimo Relative case is neither a kind of ergative nor a kind of genitive

- but there seem to be many-to-many relationships:

- every language-specific category instantiates a multiplicity of comparative concepts
- a comparative concept can be instantiated multiple times in a single language



- analysts of particular languages do not depend on the results of typology
- typologists do not directly compare the (theoretical) results obtained by specialists of particular languages
(but both contribute to the same enterprise)

5. Interim summary

- two approaches in typology (multivariate analysis; Canonical Typology) and the language-specific vs. cross-linguistic distinction
 - the list-approach (criteria/variables and values) is compatible with this distinction
 - the two approaches do not say much about the form/content dichotomy
 - the two approaches seem to assume that there is a direct relationship between language-specific categories and comparative concepts: the former are based on the latter
 - the complete list of variables/criteria (and perhaps also canonical instances) are used for cross-linguistic comparison, so they seem to correspond to comparative concepts
 - specific values of variables/instances of constructions operate on the language-specific level and can thus perhaps be interpreted as language-specific categories

(I) Multivariate analysis

- unclear what status the variables have, but probably they correspond to comparative concepts

- specific values of variables can maybe be compared to language-specific categories
- relationships between clusters are perhaps higher-order comparative concepts that are grounded in actual significant frequencies

(II) Canonical typology

- canonical instances or the list of criteria can be interpreted as comparative concepts
- it is unclear if language-specific categories are specific instances of comparative concepts (since in the ideal case a specific construction might be canonical and thus fit the comparative concept)

6. Concluding remarks

- we cannot escape opportunism

Where do the variables/criteria/canonical instances/concepts/categories come from?

- no external source or empirical methods as controlling instance
- we have to choose them and we have to define them
- therefore, they cannot be right but only useful, clear, simple, ...

Bottom-up or top-down?

- bottom-up: how many languages are enough?
- top-down: what about unfamiliar variables/criteria/canonical instances/concepts/categories and can we detect new ones?

What about frequency?

- should it play a role or not?

What about the relationship between field work and typology?

- are language-specific descriptions really independent of cross-linguistic comparison?

What about the form vs. function/semantics distinction?

- How much form (general-formal concepts) do we need in typology? (cf. Newmeyer 2007)

References

- Bickel, Balthasar. 2010. Capturing particulars and universals in clause linkage: a multivariate analysis. In Brill, I. (ed.) *Clause-hierarchy and clause-linking: The syntax and pragmatics interface*. Amsterdam: Benjamins, 51-101.
- Bickel, B. & J. Nichols. 2002. Autotypologizing databases and their use in fieldwork. In Austin, P; Dry, H; Wittenburg, P. (eds.) *Proceedings of the international LREC workshop on resources and tools in field linguistics*, Las Palmas, 26 - 27 May 2002.
- Bickel, B., T. Zakharko, L. Bierkandt & A. Witzlack-Makarevich. 2014. Semantic role clustering: An empirical assessment of semantic role types in non-default case assignment. *Studies in Language* 38, 485-511.
- Brown, Dunstan & Marina Chumakina. 2013. Introduction. In Brown, Chumakina & Corbett (eds.), 1-19.
- Brown, Dunstan, Marina Chumakina & Greville G. Corbett (eds.) 2013. *Canonical morphology and syntax*. Oxford: Oxford University Press.
- Corbett, Greville G. 2005. The canonical approach in typology. In Zygmunt Frajzyngier, Adam Hodges & David S. Rood (eds.) *Linguistic diversity and language theories*. Amsterdam: Benjamins, 25-49.
- Haspelmath, Martin. 2010. Comparative concepts and descriptive categories in cross-linguistic studies. *Language* 86, 663-687.
- Newmeyer, Frederick J. 2007. Crosslinguistic formal categories. *Linguistic Typology* 11, 133-157.