Are categories universal?

MARTIN HASPELMATH

1. Introduction

Presupposition No. 1 (Aristotelian):
Languages cannot be used or described without grammatical/syntactic categories, i.e. both speakers and linguists need categories (such as *adjective, prefix, dative, passive*).

Presupposition No. 2 (Boasian):
Latin, or English, or the latest version of GB/LFG/RRG/LDG/BLT/... are not necessarily the best model of universal grammar. Hence, descriptivists (= grammar authors) normally avoid Eurocentrism etc. and describe languages in a bottom-up fashion, "in its own terms" (and rightly so).

> **(Terminology:)**
>
> **descriptive category** (= "language-specific category"): a category that is claimed to be useful for (or to exist in) a particular language, without making any claims about other languages
>
> **universal category**: a category that is claimed to exist in all languages (*universally instantiated category*) or that could exist in any language (*universally available category;* ="cross-linguistic category").

Observation No. 1 (Dryerian):
Criteria for defining categories differ across languages. (E.g. nouns in English inflect for number, nouns in Choctaw inflect for possessor, nouns in Vietnamese do not inflect at all, ...) Hence, categories cannot be equated across languages. (Dryer 1997)

• Consequence No. 1: Descriptivists have to create their own descriptive categories.
• Consequence No. 2: Category-assignment controversies ("Is phenomenon X in my language a Y or a Z?") are pointless. (Haspelmath 2007)

> Rather than asking "Is this form an X or a Y? (e.g. a noun or a verb, an NP or a VP, and so forth)", the more appropriate question to ask is "What are the significant syntactic patterns in the language, and what are the categories that must be posited in order to enable the necessary generalisations to be stated?" (Gil 2000:174)

Observation No. 2 (Greenbergian):
Language typology is possible, i.e. apparently gives fruitful results.

2. Problems

> **Question:**
> Why is language typology possible, i.e. why are languages comparable, if they all have different categories?

Matthews 1997:199:

> "To ask whether a language 'has' some category is...to ask a fairly sophisticated question... Such warnings were once commonplace... [but] in many typological studies, scholars seem to proceed as if they were irrelevant. Cross-linguistic comparison is, on the contrary, initiated independently or in advance of detailed analyses of individual systems. One approaches each language with, in effect, a checklist....
>
> In principle, **comparisons must be based on analyses of particular systems:** but, in so many interesting cases, such analyses are lacking."

Newmeyer 1998:337-338:

> "**[F]ormal analysis of language is a logical and temporal prerequisite to language typology**. That is, if one's goal is to describe and explain the typological distribution of linguistic elements, then one's first task should be to develop a formal theory...
>
> To read the literature of the functional-typological approach, one gets the impression that the task of identifying the grammatical elements in a particular language is considered to be fairly trivial...
>
> "Assigning category membership is often no easy task... Is INFL the head of the category S, thus transforming the latter into an IP? ... Is every NP dominated by a DP? ... There are no settled answers to these questions. Given the fact that we are unsure precisely what the inventory of categories for any language is, it is clearly premature to make sweeping claims about their semantic or discourse roots. Yet much functionalist-based typological work does just that."

> **Possible answer 1:** (denies Presupposition No. 2, or Observation No. 2)
> Language typology is possible to the extent that we have figured out the universal categories.

Newmeyer 1998:342

> "...linguistic theory itself provides the means for identification [of categories across languages], though not necessarily one that can be read directly off the surface... the subject is that element underlyingly occupying the highest argument position within VP."

3. Proposed solution

> **Possible answer 2:** (favoured by me)
> Language typology is possible because typologists do not compare languages in terms of *linguistic categories*, but in terms of typologists' *comparative concepts*.

(Terminology)

> **comparative concept**: a concept that is created by a comparative linguist (=typologist) for the purpose of comparing languages and formulating cross-linguistic generalizations.

Examples of comparative concepts and generalizations based on them:

(1) *ergative case* (= a case of A, when A ≠ S, P)

  Generalization: An ergative case always has an overt marker (Dixon 1979).

(2) *subject* (= the argument with the greatest number of "subject properties", Keenan 1976)

  Generalization: The subject normally precedes the object (Greenberg 1963).

(3) *future tense* (= a tense form that has 'future time reference, E > S' as a prominent meaning)

  Generalization: Future tenses tend to be more analytic than past tenses (Dahl 1985).

(4) *reflexive pronoun* (= a pronoun that is used for local coreference, but not for local disjoint reference)

  Generalization: Reflexive pronouns are at least as long as disjoint-reference pronouns (Faltz 1985, Haspelmath 2005).

(5) *wh-movement* (= movement of a *wh*-word to a peripheral position in the clause)

  Generalization: *Wh*-movement is always to the left.

(6) *wh-word* (= a word that is used (possibly among other uses) in parametric questions (= special questions, content questions) to represent the questioned content

(7) *affix* (= a morpheme that can never occur on its own)

  Generalization: Affixes show a strong tendency to be postposed (as "suffixes") rather than preposed (as "prefixes").

Comparative concepts cannot be defined in terms of descriptive concepts (because these are language-specific), so they must be defined in terms of semantic concepts, other comparative concepts, and/or highly general relational concepts ("precedes", "is part of", "expresses", etc.).

Comparative concepts may, but need not, correspond closely to linguistic categories:

ergative case: Defined by use in A function, but in individual languages it may have a broader use (A + possessor: Eskimo *Relative case*; A + instrument, etc.)

reflexive pronoun: Defined by possibility of local coreference and impossibility of local disjoint reference, but in different languages the conditions of use may be quite different (e.g. English *himself* vs. German *sich*, cf. Reinhart & Reuland 1993)

wh-word: Defined by use in content questions, but in individual languages it may have a broader use ('who, someone', 'what, something', cf. Durie's *epistememe*, Wierzbicka's *ignorative*)

wh-movement: Defined with respect to *wh*-words, i.e. question words, but the same movement operation may also be present in other constructions such as relative clauses or comparative clauses.

Thus, the identification of elements for comparison ("comparanda") is independent of the language-particular system, pace Matthews 1997 and Newmeyer 1998.

4. Criticism from Newmeyer 2007 and Matthews 1997

Newmeyer 2007 defends universal formal categories against the claim (Dryer, Croft, Haspelmath, Lazard) that only descriptive categories and semantic categories exist, and that these are sufficient for typological generalizations.

Newmeyer observes that typological generalizations have been made about form-based categories such as "negative auxiliary" (vs. "negative particle"), "tense-aspect auxiliary" (vs. "tense-aspect particle") (Dryer 1992). Thus, purely semantic identification of comparanda across languages is not sufficient.

However, Newmeyer ignores the possibility of **category-independent comparison**, i.e. comparison based on **comparative concepts**.

To the extent that comparative concepts are (by definition) universally applicable, they can be said to be similar to universal categories. But crucially, positing a comparative concept does not commit the typologist to the claim that any language "instantiates" the comparative concept. At least some languages must have *counterpart structures*, but these need not be coherent categories within the language itself. (E.g. identifying *subjects* across languages by Keenan's definition is compatible with the view that no language has a subject category; or identifying *wh-movement* in a language is compatible with the view that the language only has a more general rule of move α)

Matthews 1997:200
> "[The typologists' approach] plainly assumes the reality of something other than the individual systems [universal grammar, or language-independent cognitive categories]. Otherwise the criteria for abstraction would be arbitrary, and critics could reasonably object that laws proposed at this level will be empty."

But if the "laws" (=universals) are interesting, this means that they are not (completely) empty. See also Lazard 2005, 2006, who explicitly recognizes that comparative categories are to some extent "arbitrary".

References

Boas, Franz. 1911. Introduction to *The Handbook of American Indian Languages*.
Dahl, Östen. 1985. *Tense and aspect systems*. Oxford: Basil Blackwell.
Dixon, R.M.W. 1979. Ergativity. *Language* 55:59-138.

Dryer, Matthew. 1992. "The Greenbergian word order correlations." *Language* 68: 81-138.

Dryer, Matthew. 1997. "Are grammatical relations universal?" In: Bybee, Joan & Haiman, John & Thompson, Sandra A. (eds.) *Essays on language function and language type.* Amsterdam: Benjamins, 115-143.

Faltz, Leonard M. 1985. *Reflexivization: A Study in Universal Syntax.* New York: Garland.

Gil, David. 2000. Syntactic categories, cross-linguistic variation and universal grammar. In: Vogel, Petra M. & Comrie, Bernard (eds.) *Approaches to the typology of word classes.* Berlin: Mouton de Gruyter, 173-216.

Greenberg, Joseph H. 1963. "Some universals of grammar with particular reference to the order of meaningful elements." In: Greenberg, Joseph H. (eds.) *Universals of grammar*, 73-113. Cambridge, Mass.: MIT Press.

Haspelmath, Martin. 2007. Pre-established categories don't exist— consequences for language description and typology. *Linguistic Typology* 11:119-132.

Keenan, Edward. 1976. "Toward a universal definition of 'subject'." In: Li, Charles N. (ed.) *Subject and topic.* New York: Academic Press, 303-333.

Lazard, Gilbert. 2005. "What are we typologists doing?" In: Frajzyngier, Zygmunt & Hodges, Adam & Rood, David S. (eds.) *Linguistic diversity and language theories.* Amsterdam: Benjamins, 1-23.

Lazard, Gilbert. 2006. *La quête des invariants interlangues: La linguistique est-elle une science?* Paris: Champion.

Matthews, Peter H. 1997. Structural linguistics in the 1990s. *Lingua* 100:193-203.

Newmeyer, Frederick J. 1998. *Language form and language function.*

Newmeyer, Frederick J. 2007. Linguistic typology requires crosslinguistic formal categories. *Linguistic Typology* 11:133-157.

Reinhart, Tanya & Eric Reuland. 1993. "Reflexivity." *Linguistic Inquiry* 24:657-720.