Linguistics and Philosophy

A Note on Descriptive Categories and Comparative Concepts in Linguistics

Klaas Willems[1]

Abstract: This note takes issue with the claim that language-specific descriptive categories and crosslinguistic comparative concepts are different "kinds of things" (Haspelmath, 2018). Against the backdrop of Eugenio Coseriu's epistemology of the language sciences it is argued that the general concepts used to make comparisons between languages are manifestations of potentially universal categories, which fall within the purview of linguistic intuition. In this sense, the relationship between categories of language and comparative concepts of linguistics reflects the creativity of language as *enérgeia*, which is a "synthetic a priori" in the Kantian sense.

Keywords: language universals; universality and generality; descriptive categories; comparative concepts; language-specific and crosslinguistic

> "Denn so wundervoll ist in der Sprache die Individualisierung innerhalb der allgemeinen Übereinstimmung, daß man ebenso richtig sagen kann, daß das ganze Menschengeschlecht nur eine Sprache, als daß jeder Mensch eine besondere besitzt." (Wilhelm von Humboldt, *Über die Verschiedenheit des menschlichen Sprachbaues*, VII: 51)[2]

---

[1] PhD, Professor, Linguistics Department, Section of General Linguistics, Ghent University, Belgium, Address: Blandijnberg 2, 9000 Ghent, Belgium, Corresponding author: klaas.willems@ugent.be.
[2] "For in language the individualization within a general conformity is so wonderful, that we may say with equal correctness that the whole human species has but one language, and that every man has one of his own."

1.1. Haspelmath (2021a) poses the question how it is possible to find a solution for what he calls "the general linguistics paradox." The paradox is, according to Haspelmath, that particular languages vary to such an extent (like other aspects of cultures) that linguists cannot study particular languages and draw conclusions about Human Language in general (Haspelmath, 2021a, p. 149). All that linguists can observe is utterances of particular languages but "the forms and rules of p[articular]-languages are not in an obvious relationship to language universals" (Haspelmath, 2021a, p. 151). Haspelmath goes on to explain that the traditional solution found in generative linguistics to solve this problem is to assume that there is a universal set of innate building blocks that can be applied to particular languages (the so-called "natural-kinds programme", in which natural kinds are akin to biological species in biology or chemical elements in chemistry), but this programme was largely given up by leading generative linguists in the 21st century. Noam Chomsky's famous (or infamous) original hypothesis of innate linguistic categories, rules and structures is nowadays no longer shared by many generative linguists.

This leaves us, according to Haspelmath, with only one alternative solution, and that is to study a wide range of languages and formulate tentative empirical language universals on the basis of wide-ranging comparisons in the vein of J. Greenberg's linguistic typology. However, the Greenbergian approach too faces a serious challenge: "worldwide comparison of languages has revealed a great diversity of categories, so that the categories of description cannot be used for comparison" (Haspelmath, 2021a, p. 27). Haspelmath concludes:

> "For the nonaprioristic comparative approach, the solution consists in recognizing that the categories of description are not the same as the yardsticks for comparison, so that language-particular studies contribute to general linguistics only in an indirect way. No linguist can simply pretend that the description of a particular language will automatically contribute to general linguistics." (Haspelmath, 2021a, p. 27)

1.2. An important question, then, is what those "yardsticks for comparison" are and how they can be found, delimited and motivated. Haspelmath claims that it is necessary to make a distinction between, on the one hand, the descriptive categories that are largely unique for every language (a "structuralist" point of view that even radicalizes the structuralist motto which says that each language should be described "in its own terms") and, on the other hand, comparative concepts created for the sole purpose of making language comparison and crosslinguistic studies possible (Haspelmath, 2010, 2018). The language-particular descriptive categories and the metalinguistic comparative concepts cannot be equated, and "the categories of one

language have no causal connection to the categories of another language" (Haspelmath, 2018, p. 85). Language-particular descriptive categories "are defined distributionally within a given language while comparative concepts are defined not distributionally but by their substantive properties" (Haspelmath, 2018, p. 96); and further:

> "Language-particular categories are defined system-internally, by other language-particular categories, but comparative concepts are defined substantively, by other comparative concepts. The distinction between system-internal categories and comparative concepts is found in the same way in other disciplines dealing with social and cultural systems and has been well-known in anthropology by the labels 'emic' (for system-internal categories) and 'etic' (for comparative concepts)." (Haspelmath, 2018, p. 109)

Thus, the central problem turns out to be that the difference between language-particular descriptive categories and metalinguistic comparative concepts is of a principled nature and seemingly insurmountable. Haspelmath maintains the view that comparative concepts and descriptive categories are entirely different "kinds of things" (Haspelmath, 2018, p. 97).

2. The issue addressed above is important, but it is often not explicitly discussed and left unresolved. In this note, I will try to cast some – admittedly dissenting – light on the issue based on E. Coseriu's account of the ways in which language-particular linguistics, general linguistics, the theory of language, the philosophy of language and the epistemology of linguistics are related to each other. Without going into details, I draw in particular on publications such as the following: Coseriu (1958), (1962), (1968), (1974), (1976), (1992), (2007), (2015), among others – at the same time also a gentle invitation *ad doctum lectorem*…

2.1. First of all, the assumption that there might be a "causal connection" between the categories of one language and the categories of another language rests on a controversial understanding of what categories in language are. There is agreement among scholars that categories are no objects in the physical world but constructs of the human "mind-cum-body." Causal relations, on the other hand, exist between objects, not between categories. According to Coseriu, language itself is not an object but a human activity (*enérgeia*), which however can be turned into an object of enquiry in different ways, to different extents and for different purposes (Coseriu,

1956; 1968). For instance, if a relation between the category of verb and the category of adjective can be ascertained, then this means that these two parts of speech are related to each other as modalities or modes of meaning, i.e. they constitute "categorial meanings" that correspond to the differences in the "mode of apprehension" ("in der Weise der Erfassung," E. Husserl) as opposed to lexical meanings, e.g.: *made of stone* (Noun) – *stone someone* (Verb) – *a stone wall* (Adjective) (Coseriu, 1974, p. 62). All meanings in language are in this sense intentional constructs of speakers, not reflections of reality or some pre-linguistic thought in the human species. Even the corresponding word 'classes', i.e. the classes of adjectives and verbs (which themselves are no classes but categories), are not causally connected, they are the product (*érgon*) of the intentional activity of speaking by generations of speakers.

2.2. Universality and generality are distinct concepts with regard to language. Any category found in a language, even if only in one language, is a possible universal category of Human Language (Coseriu, 1968, p. 70; Coseriu, 1974, pp. 49-50). By contrast, only empirical research of particular languages and their comparison can determine whether, or to what extent, a specific category is general in the languages that are being studied. This distinction is important in order to distinguish between the (possibly diverse) manifestations of a single category and an aprioristic generalization. The latter is addressed by Haspelmath as follows:

> "There are usually enough similarities between different languages to make it tempting (and in some sense useful) to reuse the same terms (e.g. 'auxiliary' both for English modal auxiliaries and for German tense auxiliaries), but the categories are really defined by their language particular structural behaviour (e.g. the lack of non-finite forms of English modal auxiliaries), not by instantiating some general (aprioristic) category" (Haspelmath, 2021a, p. 22).

From a Coserian point of view, to assert essential universals, whether based on intuition or on "intuition-cum-practice" (empirical research), is not aprioristic, but generalizations can be aprioristic. For instance, if Sapir (1921) is right in stating that in Nootka parts of speech such as verb and noun are only differentiated in actual discourse through specific morphology but that the lexical items by themselves are not marked for either of the parts of speech, then this means that Nootka does not have verbs and nouns as word classes. It would be an aprioristic generalization based, e.g., on the study of Indo-European languages to claim that Nootka has verbs and nouns as encoded classes of the lexicon. However, it would not be erroneous, nor

would it be aprioristic, to claim that in Nootka, too, verbs and nouns exist, because Nootka, like any other language, distinguishes between the categorial semantic functions verb and noun, and necessarily so (Coseriu, 1974, p. 51). By contrast, it is well-known that certain languages are said not to distinguish between verb and adjective or between noun and adjective. If it can be shown that these languages not only lack an adjective class but also the categorial semantic function of adjective in discourse, then it would be an aprioristic claim to maintain that they have adjectives (cf. Dixon, 2004, p. 12, who once maintained that there are languages without adjectives but later revised his opinion, asserting that all known languages can be shown to have adjectives).[1]

2.3.1. The comparison of categories in different languages by means of comparative concepts presupposes the knowledge and the (possibly only intuitive and incomplete) definition of the categories, not the reverse. It would be circular to think that this knowledge or this definition is the outcome of the comparison, even though the comparison, or any empirical research, may modify, enrich and precisify the preliminary definition one adopts and/or the theoretical notion. This is actually something that occurs all the time, in particular in the initial stages of a science or a specific subdiscipline. A very general example can serve as an illustration. Take the issue of "what language is," i.e. the definition of Human Language. Linguists have for a long time thought that natural language was necessarily dependent on the oral-aural modality for its material realization until it dawned on them, particularly since the 1960s, that natural languages do not have to be realized by means of sounds. The manual-visual modality of sign language is as good a resource of signifiers than sounds. Ever since, language has been defined in such a way that the modality in which signifiers are realized to produce the bilateral signs of natural language ("signifier + signified") is either left unspecified or, alternatively, differentiated between spoken languages (oral-aural modality) and sign languages (manual-visual modality). It might seem that this new definition is simply the outcome of empirical research. But this is not accurate. It is possible that many scholars had to revise some of their basic views, but native sign language users and bilingual speakers/signers (e.g., in the broadly bilingual community of Martha's Vineyard in the 19th and early 20th century) knew it all along: signifiers of natural language are not necessarily sounds. Hence, the definition may have been wrong in the heads of many individuals,

---

[1] Note, however, that Dixon (2004) consistently refers to adjectives as a separate "class of words," not to the categorial semantic function "adjective." He can thus agree, e.g., with Kruspe (2004) on Semelai and consider adjectives "as a well-defined sub-class of verbs" (Dixon, 2004, p. 12).

prompting the clarification that eventually turned out to be necessary, but the "notion of language" itself – the "Sprachidee" invoked by Wilhelm von Humboldt[1] (cf. Coseriu, 2015, II, p. 12), which is apprehended intuitively by human beings – is a priori such that it provides room, i.e. leeway, for any adjustment that a fully developed definition requires. Importantly, "a priori" in this context does not refer to an innate idea but is meant in a Kantian sense: while experience starts off with an observable, empirical fact or object, the source of the structuring of that experience lies within the mental capabilities of human beings, which are a matter of consciousness and creativity.

Applied to linguistics, this means that the relationship between the universal and the general reflects the relationship between the universal and the general in language itself: language is an original "Synthese *a priori*" (Coseriu, 2015, II, p. 403). This is a consequence of the fact that language is incessantly created anew in acts of speaking and writing, based on the knowledge of languages ("traditional systems") that people share – a knowledge that is itself always in flux (see Coseriu, 2007, p. 70). This creativity entails that knowledge of language cannot be reduced to accumulating abstractions ("generalizations") over usage events or simply be equated with knowledge of conventionalized resources.

It therefore seems that no coherent account of the relationship between potentially universal categories and comparative concepts in linguistics is possible unless Kant's key conceptual distinction between the "Anfang" and the real "Ursprung" or "Quelle" of knowledge is observed:

> „Die Synthesis eines Mannigfaltigen aber (es sei empirisch oder a priori gegeben) bringt zuerst eine Erkenntnis hervor, die zwar anfänglich noch roh und verworren sein kann, und also der Analysis bedarf; allein die Synthesis ist doch dasjenige, was eigentlich die Elemente zu Erkenntnissen sammlet, und zu einem gewissen Inhalte vereinigt; sie ist also das erste, worauf wir Acht zu geben haben, wenn wir über den ersten Ursprung unserer Erkenntnis urteilen wollen" (Kant, *Kritik der reinen Vernunft*, A77/B103).[2]

It also bears pointing out that the preliminary knowledge of a category ("notional knowledge") and a "definition" that is established as a result of empirical

---

[1] Humboldt believed that sounds coming from the human mouth and perceived by the human ear constitute the ideal signifier of natural languages (cf. Trabant, 1990, p. 8).

[2] "The synthesis of a manifold, however, (whether it be given empirically or *a priori*) first brings forth a cognition, which to be sure may initially still be raw and confused, and thus in need of analysis; yet the synthesis alone is that which properly collects the elements for cognitions and unifies them into a certain content; it is therefore the first thing to which we have to attend if we wish to judge about the first origin of our cognition" (Kant, *Critique of pure reason*, A77/B103).

comparative research, are not identical. Haspelmath writes:

> "There is an open-ended range of phenomena that one might take into account in formulating a generalization, and we are often biased toward particular conclusions […]. Again, this does not mean that we should not be inspired by what other languages do, but we cannot simply assume that different languages have the same categories when they are manifested differently" (Haspelmath, 2021b, p. 145).

But regardless whether you assume or not that "different languages have the same categories when they are manifested differently," in either case you necessarily presuppose the existence of the categories in the different languages, otherwise you could not determine whether they are "manifested differently" in different languages. This shows three things:

• the manifestation of a linguistic category is not the logically primary datum for the language scientist; it is what it says, a manifestation of something, and that something, the category, is logically prior to its manifestation(s); for instance, the knowledge of the part of speech "verb" precedes the observation of the existence of the verb class (and any other word class), its formal and semantic properties etc. in a particular language;

• manifestations of a linguistic category are not compared directly but on the basis of the preliminary knowledge of the category, which is the indispensable *tertium comparationis*: in other words, you need to know what verbs, nouns and adjectives (i.e., their categorial meanings) are if you want to compare their manifestations in different languages; likewise, you need to know what an article, an affix, a specific word formation pattern, a passive, a ditransitive argument construction etc. are before you can ask yourself whether a particular language actually has an article, an affix, a specific word formation pattern, a passive, a ditransitive argument construction etc.

• linguistic research has shown that there is a manifold of categories in the languages of the world, and many of them are surprising to students of the far better researched languages (for examples, see e.g. the Rara & Universals Archive: "Das grammatische Raritätenkabinett"[1]); but whenever a new, uncommon or otherwise remarkable "rarum" is discovered, it is accommodated on the basis of an already apprehended category, by aligning it with an existing one, narrowing down or broadening a category, etc.; the purview of linguistic intuition is not measurable, nor are its 'traits' finitely enumerable.

---

[1] https://typo.uni-konstanz.de/rara/category/raritaetenkabinett/

2.3.2. The directedness of the relationship between a category and its manifestation(s) pervades, mutatis mutandis, linguistic analysis, also with regard to closely related languages. To give just one example, consider L. Hjelmslev's well-known comparison of lexical items in three languages with regard to the 'zone of purport' (Hjelmslev, 1961, p. 54) that is covered by Danish *træ* and *skov*, German *Baum*, *Holz*, *Wald* and French *arbre*, *bois*, *forêt*. In a number of more recent studies, the comparison has been extended with Spanish and some of the contrastive relations have been rearranged (see Cigana & Polis, 2023, for discussion). This resulted in a representation as shown below:

	TREE	WOOD (STUFF)	FIREWOOD	SMALL FOREST	LARGE FOREST
German	Baum	Holz		Wald	
Danish	træ			skov	
French	arbre	bois			forêt
Spanish	árbol	madera	leña	bosque	selva

It is tempting to think of the top row in the table as representing the "meanings" of the words in the languages that are being compared. However, this would be inaccurate. The top row contains the "labels" of general denotational functions that are rendered by means of the English language, which is here used as a metalanguage (Cigana & Polis, 2023, p. 96). These labels are identified and established by dint of the signifieds of words, which constitute a – potentially universal – category of meaning encoding in natural language. The metalinguistic labels in the top row can be considered more or less adequate paraphrases of the signifieds of the words *árbol*, *madera*, *leña*, *bosque* and *selva* in Spanish, and the same holds, e.g., for the German word *Baum* and the French word *arbre* with regard to TREE. However, what the comparison above all shows is that the lexical categories in the four languages are different. These categorial differences would be obfuscated if, e.g., the correspondance between the three general denotational functions TREE + WOOD (STUFF) + FIREWOOD in the metalanguage and the semantic scope of the word *træ* were construed as representing the "meaning" of *træ* in Danish. Rather, TREE + WOOD (STUFF) + FIREWOOD collates generalized abstractions over usages of the word *træ* (likewise, SMALL FOREST + LARGE FOREST collates generalized abstractions over usages of the word *skov*). But unlike in Spanish, German and French, there is no word in Danish with the signified TREE; the word *træ* has an underspecified signified which can be used to denote, among other things, a tree. Thus, not only should we observe the difference between the "multifunctionality" of a single word in terms of general denotational functions and its unitary signified, but

it is the signified of a word that is the categorial precondition for the manifestation of its denotational multifunctionality, not the other way round.

3. In summary, it is important that the "universal" in language is not mixed up with the "general," and that the notion of "universals" in language is accordingly differentiated. Coseriu's (1974) distinction between empirical, essential (rational) and possible universals (and the two additional hybrid selective and implicational universals; cf. Willems, 2016) provides an appropriate basis, and an explicit rationale, for addressing the – apparently insurmountable – paradox of the difference between language-particular categories and comparative crosslinguistic concepts.

References

Cigana, L. & Polis, S. (2023). Hjelmslev, a forerunner of the semantic maps method in linguistic typology? *Acta Linguistica Hafniensia*, 55(1), 93-118.

Coseriu, E. (1956). *La creación metafórica en el lenguaje*. Montevideo: Universidad de la República, Facultad de Humanidades y Ciencias, Instituto de Filología, Departamento de Lingüística. German translation: Die Metaphernschöpfung in der Sprache, in: E. Coseriu, *Sprache. Strukturen und Funktionen* (1979). Tübingen: Gunter Narr, 15-44.

Coseriu, E. (1958). *Sincronía, diacronía e historia. El problema del cambio lingüístico*. Montevideo: Universidad de la República, Facultad de Humanidades y Ciencias, Instituto de Filología, Departamento de Lingüística. German translation: *Synchronie, Diachronie und Geschichte. Das Problem des Sprachwandels* (1974). München: Wilhelm Fink.

Coseriu, E. (1962). *Teoría del lenguaje y lingüística general. Cinco estudios*. Madrid: Gredos. 2nd edition: Madrid: Gredos, 1967, reprint: Madrid: Gredos, 1970. German translation: *Sprachtheorie und allgemeine Sprachwissenschaft. 5 Studien* (1975). München: Wilhelm Fink.

Coseriu, E. (1968). Der Mensch und seine Sprache. H. Haag & F. P. Möhres (eds.), *Ursprung und Wesen des Menschen*, pp. 67-79. Tübingen: J. C. B. Mohr (Paul Siebeck).

Coseriu, E. (1974). Les universaux linguistiques (et les autres). Luigi Heilmann (ed.), *Proceedings of the Eleventh International Congress of Linguists*, I. Bologna: Società editrice il Mulino, pp. 47-73. English translation: Linguistic (and other) Universals, in: A. Makkai, V. B. Makkai & L. Heilmann (eds.) (1977), *Linguistics at the Crossroads*. Padova and Lake Bluff: Testi e Studi, III, pp. 317-346.

Coseriu, E. (1976). Logique du langage et logique de la grammaire. J. David & R. Martin (eds.), *Modèles logiques et niveaux d'analyse linguistique*, pp. 15-33. Paris: Centre d'Analyse syntaxique de l'Université de Metz.

Coseriu, E. (1992). *Einführung in die Allgemeine Sprachwissenschaft*, 2nd edition. Tübingen: Francke.

Coseriu, E. (2007). *Sprachkompetenz. Grundzüge der Theorie des Sprechens*, 2nd revised edition.

Tübingen: Narr.

Coseriu, E. (2015). *Geschichte der Sprachphilosophie*, vol. I: *Von Heraklit bis Rousseau*, vol. II: *Von Herder bis Humboldt*. Tübingen: Narr.

Dixon, R. M. W. (2004). Adjective classes in typological perspective. R. M. W. Dixon, & A. Y. Aikhenvald (Eds.), *Adjective classes: A cross-linguistic typology*, pp. 1-49. Oxford: Oxford University Press.

Haspelmath, M. (2010). Comparative concepts and descriptive categories in crosslinguistic studies. *Language*, 86(3), 663-687.

Haspelmath, M. (2018). How comparative concepts and descriptive linguistic categories are different. D. Van Olmen, T. Mortelmans & B. Frank (eds.), *Aspects of linguistic variation: Studies in honor of Johan van der Auwera*, pp. 83-113. Berlin: De Gruyter Mouton.

Haspelmath, M. (2021a). General linguistics must be based on universals (or non-conventional aspects of language). *Theoretical Linguistics*, 47(1-2), 1-31.

Haspelmath, M. (2021b). How to tear down the walls that separate linguists: continuing the quest for clarity about general linguistics. *Theoretical Linguistics*, 47(1-2), 137-154.

Hjelmslev, L. (1961). *Prolegomena to a theory of language*, 2nd, revised edition. Madison: University of Wisconsin Press. Danish original: *Omkring sprogteoriens grundlæggelse* (1943). København: Munksgaard.

Humboldt, W. von (1836/1998). *Über die Verschiedenheit des menschlichen Sprachbaues und ihren Einfluß auf die geistige Entwicklung des Menschengeschlechts*. Edited by Donatella Di Cesare, Paderborn, etc.: Ferdinand Schöningh. English translation: *On language: The diversity of human language structure and its influence on the mental development of mankind* (1988). Cambridge: Cambridge University Press.

Kant, I. (1787/1956): *Kritik der reinen Vernunft*. Edited by Raymund Schmidt, Hamburg: Felix Meiner. English translation: *Critique of pure reason* (1998), translated and edited by Paul Guyer and Allen W. Wood. Cambridge: Cambridge University Press.

Kruspe, N. (2004). Adjectives in Semelai. R. M. W. Dixon, & A. Y. Aikhenvald (Eds.), *Adjective classes: A cross-linguistic typology*, pp. 283-305. Oxford: Oxford University Press.

Sapir, E. (1921). *Language: An introduction to the study of speech*. New York: Harcourt, Brace & World.

Trabant, J. (1990). *Traditionen Humboldts*. Frankfurt am Main: Suhrkamp.

Willems, K. (2016). The universality of categories and meaning: a Coserian perspective. *Acta Linguistica Hafniensia*, 48(1), 110-133.

Rara & Universals Archive. https://typo.uni-konstanz.de/rara/.