

The Role of Epistemology and Ontology in Terminology

Johann-Mattis List
Chair for Multilingual Computational Linguistics
University of Passau

Terminology is essential for scientific research. Depending on its purpose, it may be diagnostic, expressing judgments made by researchers, or descriptive, referring to empirical observations. Recent discussions about the popular colexification term have revealed that scholars use two interpretations, one narrow one, in which colexification is seen as a cover term for polysemy and vagueness, and one broad one, in which colexification is seen as a cover term for polysemy, vagueness, and homonymy. By discussing the role that the epistemological and the ontological perspective play in the context of scientific constructs and scientific terminology, I argue that colexification is a diagnostic term in the narrow definition, while it is a descriptive term in the broad notion. From an ontological perspective, the narrow definition can be seen as important, as it distinguishes processes of semantic change from processes of sound change. From an epistemological perspective, however, the narrow definition has difficulties to be applied in practice. I conclude that both kinds of terminology and both perspectives are useful for scientific research.

1 Introduction

Science deals in a very simplified notion with the idea that particular phenomena that we consider as interesting can be explained in models that consists of *objects* and *processes*. Objects represent constructs that we consider as discrete. Processes represent *constructs* that relate objects in certain ways, they can thus be described as relations between objects.

Constructs themselves can be seen as “conceptual term[s] used to describe a phenomenon of interest” (Edwards and Bagozzi 2000, 156f). Constructs can be seen as some placeholder that we use to refer to the phenomena. By defining constructs, we reflect that we consider these phenomena to be “real” in the sense of being measurable through scientific studies, but we also admit – as we should always do in science – that our understanding of these phenomena is not complete, and that we approach them in scientific models that do not capture the phenomena in all their complexity.

[...] although constructs are terms researchers literally construct, or put together [...], we intend that constructs refer to phenomena that are real and exist apart from the awareness and interpretation of the researcher and the persons under study. (Edwards and Bagozzi 2000, 157)

If we consider *inflation* as a scientific construct, we can say that we take it for granted that prices for goods can rise in societies that employ monetary systems to exchange goods. How to measure inflation in due detail, however, is a secondary question that may yield controversial discussions among scientists and the broader public (List 2025).

Constructs are quite important in linguistics as well, specifically in the field of *linguistic typology*, where scholars tend to compare specific properties of languages that are controversially disputed among experts. While some consider them as difficult to define, others may even dispute if they exist after all. Haspelmath (2010) has introduced the term *comparative concept* in order to refer to these *tertia comparationis*. Comparing the discussion about scientific constructs in other branches of science, such as psychology, we can probably say that the notion of the construct comes very close to the notion of the comparative concept. The latter may, however, be seen as narrower in its scope, given that it is specifically relevant for linguistic typology and cross-linguistic studies.

No matter which term we chose, be it *construct* or *comparative concept*, it seems useful to distinguish these supposed phenomena (objects and processes) that we plan to investigate in our research from the *real* objects and processes that they reflect.

2 Epistemology and Ontology

When dealing with scientific constructs, we can distinguish two fundamental perspectives in science, the *ontological perspective* and the *epistemological perspective*. The former reflects what we think is the case with respect to the world. The latter reflects what we think we can *know* about this supposed truth.

It is clear that what we think is true differs from what *is* true, and it is also clear that the term *ontology* may also be used to refer to what *is* true independently of what we think is true. However, when working scientifically, we can never really tell whether something holds for real. We will always be left with some doubts. These doubts may be very small, but depending on the topic we deal with, they may also be so huge that we would not be surprised if they are confirmed in future work. As a result, when I talk about the ontological perspective, I follow Huber (2014) in treating it as something close to our initial *convictions*, our basic *model* that holds for some phenomenon that we want to investigate, while the epistemological perspective asks how well we can actually find evidence for these convictions in the real world.

Thus, following Huber (2014), one can say that an ontological question regarding scientific objects or processes would ask if they exist, if they are real, or what defines them. An epistemological question, on the other hand, would ask if we can identify them properly from other objects and processes, and to what degree we can be certain to be dealing with these objects and processes in a given situation.

3 Diagnostic and Descriptive Terminology

Terminology in science fulfils different functions, depending on what the terms we create in the different scientific fields refer to. In general, terms are needed to *name* our scientific constructs, that is, objects and processes, and phenomena, that we want to explain scientifically. But since terms are never free of our general assumptions about the phenomena we want to describe, they are always related to our particular *ontologies*, that is, our *theories* or *models*, which may differ from school to school and even from scholar to scholar.

When terminology is used to tell certain objects or processes apart, the assignment of a term to an object or process reflects a *judgment* by the researcher. In such cases, our scientific terminology is similar to the diagnosis of a doctor who tries to identify a certain disease (which is again a scientific construct). In Jacques and List (2019) we called such uses of terminology *explanatory* and distinguish them from *descriptive terminology* (Jacques and List 2019, 141). Our example was the term *shared innovation* in phylogenetic reconstruction, which is often falsely treated as a mere observation, although its use always entails a clear *diagnosis*, since it must be distinguished from a *shared retention* and cases of *parallel evolution* or *lateral transfer*. Instead of the term *explanatory terminology*, one might also use the term *diagnostic terminology*, which has, however, a specific meaning in medicine.

In any case, it seems important to keep in mind, when dealing with terminology, that terms are not always merely descriptive, describing an observation, but can also come along with explicit judgments made by the researchers. In general, it seems useful to distinguish terms that require some kind of diagnosis from terms that reflect an observation, even if the distinction between diagnosis and observation may not always be a straightforward one. We need *descriptive terms*, since our *diagnostic terms* can easily exhibit epistemological conflicts. That is, we may often encounter situations where a proper judgment cannot be made, and where it would instead be preferable to embrace our own uncertainty.

4 Results

The term *colexification* is an example for the power of terminology over scientific research. Ever since the term was first introduced by François (2008), it has inspired numerous theoretical and also empirical studies dealing both with the term itself and with its application to linguistic data. In the original definition given by François, the term refers to cases where a particular lexical form in one languages refers to more than one *sense*.

A given language is said to colexify two functionally distinct senses if, and only if, it can associate them with the same lexical form. (François 2008, 170)

The study leaves it open, how a lexical form should be defined, and if one should allow for two lexical forms with the same pronunciation in one and the same language or not. Later remarks in open discussions by François (List et al. 2026), however, as well as certain examples provided in the original text from 2008 make it clear – at least in the retrospective – that François was thinking of *colexification* as a cover term for *polysemy* and *vagueness*, rather than – as I myself and several people with whom I have collaborated in the past have assumed for quite some time – a cover term for *polysemy*, *homophony*, and *vagueness* (see Tjuka 2026, 5).

While my own misinterpretation of François' terminology may be seen as a classical example of scientists who do not pay enough attention to what they read, the very fact that not only I, but also quite a few people with whom I have collaborated in the past, have misinterpreted the scope of the term, has led to concrete progress in quantitative approaches to lexical typology that would *not* have been possible otherwise.

The progress itself touches the above-mentioned distinction of an ontological from an epistemological perspective as well as the importance of *descriptive* as opposed to *explanative* or *diagnostic* terminology. The crucial misunderstanding underlying our “broad” definition of colexification as including homophony, vagueness, *and* polysemy

can be seen as a misunderstanding in which a *diagnostic term* was interpreted as a *descriptive term*.

Haspelmath (2026) summarizes the discussion about the definition underlying the term *colexification* nicely, in emphasizing that colexification can be seen as a *comparative concept* or a *scientific construct* and that it refers linguistically to a perspective where one compares phenomena *across* languages (“g-linguistics”), rather than investigating phenomena *inside* one and the same language (“p-linguistics,” see Haspelmath 2019). In order to establish a certain colexification of two senses (with senses actually referring to individual *comparative concepts*) in a given language variety, one must check, if the language uses one lexical form to express both concepts, or if it uses two separate forms, each expressing one of the two concepts. The two definitions of colexification, the narrow one and the broad one, differ in how they define a lexical form. The narrow definition allows for two lexical forms to have the same shape in their sound sequences (they sound identical), while the broad definition does not allow to postulate two forms if both have the same phonetic shape.

The narrow definition can be seen as *diagnostic*, since we need to make a concrete *diagnosis* in order to identify two forms where we have only one sound sequence in a given language. The broad definition is *descriptive*, since it does not try to distinguish polysemy from homophony. Haspelmath (2026) emphasises the importance of allowing for the distinction of multiple lexical forms in one language, even if they have the same shape, but he also seems to agree that the distinction can be quite difficult in practice. While it may be useful to establish a cover term for polysemy and vagueness from an ontological perspective, it seems difficult to apply this term to actual examples from an epistemological perspective.

In this context, it may be important to emphasize that the term *polysemy* itself was coined as a *diagnostic* term, pointing to those cases where a form acquired more meanings over time due to processes of semantic change (Bréal 1897). If we identify a case where different concepts are expressed by different sound sequences as *polysemy*, we thus provide a particular *diagnosis* or a *judgment* on the history of the form and its senses. Whenever this diagnosis cannot be made, we are in a situation where the ontological perspective – we know one lexical form can acquire more senses over time but we also know two lexical forms that sound differently can come to sound alike at some point in time due to sound change – is in conflict with the epistemological perspective – we do not know always what process lead to the broad colexification.

Being able to distinguish these two perspectives seems urgent and important to me. The success of the *Database of Cross-Linguistic Colexifications* (<https://clics.clld.org>, Rzymiski et al. 2020; Tjuka et al. 2025) as a database in which *potential narrow*

colexifications are automatically assembled from *clear broad* colexifications also confirms this need. Without the misinterpretation of *colexification* as a term that does not seek a diagnosis, we might have never started to create colexification networks from cross-linguistic data in the first place (List et al. 2013).

5 Conclusion

In this short overview, I have tried to argue that the distinction between the two notions of colexification is one between diagnostic and descriptive terminology and between different perspectives on research. François' narrow notion of colexification is diagnostic in so far as it requires an explicit judgment by the researcher in order to determine if a give sound sequence expressing two or more independently defined concepts in one particular language reflects one or more lexical forms. It maintains an ontological perspective in so far as it points to a crucial difference with respect to the processes underlying polysemy and vagueness on the one hand and homophony on the other hand. The broad notion that we used in previous studies, on the other hand, sacrifices the precision with respect to the known processes of semantic change and sound change in favor of a terminology that is more in line with what we *can* know without having deeper access to the history of individual languages. The broad colexification term is thus descriptive in nature and more precise with respect to what we know, but one could argue that it conflicts with our deeper knowledge of the processes underlying polysemy, homonymy, and vagueness.

That the broad, descriptive notion can be useful has been demonstrated by the *Database of Cross-Linguistic Colexifications* as well as the more recently published *Database of Cross-Linguistic Partial Colexifications* (<https://clips.digling.org>, Bocklage et al. 2026). Here, we start from the description of cases of broad colexification and later extend it with an analysis that helps us to provide a diagnosis for the most frequently recurring colexification patterns identified in the data. All in all, it seems that both perspectives are important. We need the ontological perspective and the explicit attempt of making a judgment regarding our research data, *and* the epistemological perspective and the explicit attempt to describe the data in its unanalyzed form.

References

- Bocklage, Katja, Thanasis Georgakopoulos, Robert Forkel, and Johann-Mattis List. 2026. "CLIPS – A database of cross-linguistic partial colexifications." In *Knowledge Commons*, No. 0, vol. 0. <https://doi.org/10.17613/cjqny-jw696>.
- Bréal, Michel. 1897. *Essai de sémantique: Science Des Significations*. Hachette.

- Edwards, Jeffrey R., and Richard P. Bagozzi. 2000. "On the nature and direction of relationships between constructs and measures." *Psychological Methods* 5 (2): 155–74. <https://doi.org/10.1037/1082-989x.5.2.155>.
- François, Alexandre. 2008. "Semantic maps and the typology of colexification: intertwining polysemous networks across languages." In *From polysemy to semantic change*, edited by Martine Vanhove. Benjamins.
- Haspelmath, Martin. 2010. "Comparative concepts and descriptive categories." *Language* 86 (3): 663–87.
- Haspelmath, Martin. 2019. "Confusing p-linguistics and g-linguistics: Philosopher Ludlow on «framework-free theory»." *Diversity Linguistics Comment* 7 (6). <https://dlc.hypotheses.org/1801>.
- Haspelmath, Martin. 2026. "Coexpression, colexification, cogrammatication: Three new terms of comparative linguistics and what they (might) mean." *Diversity Linguistics Comment* 2026 (8). <https://doi.org/10.58079/16o10>.
- Huber, Hans Dieter. 2014. "Ontologische und epistemologische Voraussetzungen ästhetischer Bildung." In *Bild und Bildung: Praxis, Reflexion, Wissen im Kontext von Kunst und Medien*, edited by Barbara Lutz-Sterzenbach. Lehmanns.
- Jacques, Guillaume, and Johann-Mattis List. 2019. "Save the trees: Why we need tree models in linguistic reconstruction (and when we should apply them)." *Journal of Historical Linguistics* 9 (1): 128–66. <https://doi.org/10.1075/jhl.17008.mat>.
- List, Johann-Mattis. 2025. "Von gefühlten Tatsachen." *Von Wörtern Und Bäumen* 9 (10). <https://doi.org/10.58079/14zc7>.
- List, Johann-Mattis, Katja Bocklage, and Thanasis Georgakopoulos. 2026. "The ambiguity of the colexification term." *Computer-Assisted Language Comparison in Practice* 9 (2): 85–93. <https://doi.org/10.15475/calcip.2026.2.2>.
- List, Johann-Mattis, Anselm Terhalle, and Matthias Urban. 2013. "Using network approaches to enhance the analysis of cross-linguistic polysemies." *Proceedings of the 10th International Conference on Computational Semantics – Short Papers*, 347–53.
- Rzyski, Christoph, Tiago Tresoldi, Simon Greenhill, et al. 2020. "The Database of Cross-Linguistic Colexifications, reproducible analysis of cross- linguistic polysemies." *Scientific Data* 7 (13): 1–12. <https://doi.org/10.1038/s41597-019-0341-x>.
- Tjuka, Annika. 2026. *Variation in word meanings: Cross-Linguistic Patterns and Causes of Colexifications*. World Histories of Lexicography and Lexicology 3. Language Science Press. <https://doi.org/10.5281/zenodo.20639912>.
- Tjuka, Annika, Robert Forkel, Christoph Rzyski, and Johann-Mattis List. 2025. "Advancing the Database of Cross-Linguistic Colexifications with New Workflows and Data." *Proceedings of the 16th International Conference on Computational Semantics*, 1–15. <https://aclanthology.org/2025.iwcs-main.1>.

Funding Information

This project was supported by the ERC Consolidator Grant ProduSemy (PI Johann-Mattis List, Grant No. 101044282, see <https://doi.org/10.3030/101044282>). Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Council Executive Agency (nor any other funding agencies involved). Neither the European Union nor the granting authority can be held responsible for them.