

From: Daniel Hieber
To: [Jean Hieber](#)
Subject: Fwd: LT
Date: Thursday, May 9, 2013 10:32:05 AM

----- Forwarded message -----

From: Plank <frans.plank@uni-konstanz.de>
Date: Tue, May 7, 2013 at 11:42 AM
Subject: LT
To: "Daniel W. Hieber" <dwhieb@gmail.com>

Dear Daniel,

this was an unusually difficult decision, and it took us unusually long to reach it. After deliberating the reviews back and forth, the opinion prevailed not to publish your flexible categorisation paper. We think it will probably be in your own interest, too, to either try somewhere else or (better still) sit back and then make a new start on this vexed question where it isn't easy to rise above controversy (even in such elementary matters as what is at issue) and substantially take things further.

As you can see from the reviews below, the reviewers (the same as the first time round, with one original reviewer defaulting) are torn between acknowledging that you've attempted something important, and have stepped up your efforts considerably, and wishing for more clarity and conciseness.

We don't usually consider resubmissions of papers where a decision is negative. Still, since part-of-speech/word-class/lexeme categorisation is such a central issue, much discussed in LT, and since the reviews do see potential in your approach, we remain interested in your work.

Points to consider, to go by the reviews, would primarily include these:

- length/conciseness;
- typological generalisations, to accompany classifications;
- connected with this, the rationale of "canonical" typology;
- What is it all about?: lexemes vs parts of speech [I'm not sure I'd go with the reviewers here; still, it should be clear what is being categorised and typologised].

I'm sorry not to have better news after such a long time. But perhaps the news aren't entirely bad, either.

With best regards
Frans

Title: A canonical typology of flexible categories
Author: Daniel W. Hieber
Received: 29 Jan 2012
Revised: 30 Dec 2012

REVIEW 1

GENERAL IMPRESSION

On the one hand, this revised version seems much improved to me in terms of

- Discussion of Canonical Typology and the place of the advocated approach to parts of speech categorization amongst other approaches;
- Addition of examples

On the other hand, the manuscript is now very long. I think that in several places the additional material is written in a very readable, almost 'think-aloud' type of style, the downside of which is that it is rather space-consuming.

Another, more serious problem is that the paper still does not seem to show how the proposed method helps us to better understand cross-linguistic/universal tendencies in systems of lexical categorization. More specifically: the "central problem" it addresses concerns "the deadlock that arises from debates about lexical flexibility" (i.e. does language X have category Y or not?) . The proposed solution is to dissect flexibility into a number of sub-phenomena, each of which a particular language may exhibit to different degrees. The final section illustrates how a language (Mundari in this case) may be typologized using this method. While it seems conceivable to compare multiple languages, each typologized along the same 11 criteria, the paper does not discuss what kinds of insights this would give into universal principles of lexical categorization – insights beyond the idea that flexibility is not a primitive notion, but rather a multi-faceted phenomenon, which in itself is not entirely novel.

In sum:

- While I agree that the "deadlock problem" is a real and serious one, I am not convinced that the proposed solution is the right one.
- While the additions to the previous versions are welcome, a final round of tightening things up again might be helpful.

MORE SPECIFIC COMMENTS

- Section 2

About section 2.1: I can see that using a fine-grained typology in which multiple parameters with multiple values are considered (and their possible co-variation can be determined) is a merit, but not one specifically tied to the Canonical method. At the same time, I don't see how the Canonical Typology approach is particularly helpful for the typologist to overcome problems of different labeling (i.e. 'not comparing like

with like'). Also, it seems that there is no independent justification for using a particular set of parameters/criteria - it remains unclear how we get to the "agreed-upon" (p. 14), "indisputable" (quote p. 15) canon in the first place. (This is different in a data-driven approach, where new parameters/values are added in account for newly-found typological variation. To some, therefore, "precategorization of logically possible realizations" would rather seem to be a disadvantage of the Canonical Approach (see p. 13, section 2.3).

N.B. These comments are meant to indicate where the argumentation could perhaps be stronger: I appreciate that there will always remain disagreements about the merits and deficiencies of specific approaches to typology.

- p.20, Table 3: I don't find the English examples in the table very helpful. Since the different cells are exemplified by specific phenomena in various languages in the text, I would suggest to leave the English examples out.

- p.21, 45, 50: In several places, it is mentioned that individual researchers may disagree about the status of certain non-canonical phenomena. It is not clear how such disagreements would be meaningfully solved – therefore, this begs the question whether they wouldn't compromise the main goal of the approach, namely to solve such disagreements...

- p. 25, 1st line: The use of the term "zero-conversion" is confusing here: Hengeveld and Rijkhoff don't 'believe' in zero-conversion. In their view, conversion is irrelevant to flexible lexemes, since they are not supposed to change category anyway; rather, they belong to a flexible category and merely have certain aspects of their semantics 'profiled' when used in a specific syntactic function.

- p.26, final part of section 3.3: Hengeveld and colleagues have proposed a number of revisions of the typology in Table 1, due to counterexamples attested in larger language samples (see Hengeveld & Van Lier 2008, 2010). Perhaps, for completeness' sake, this recent work should be mentioned too.

- p. 32, above example (17): Tagalog is not an exclusively isolating language (and anyway this term seems to be used in the old-fashioned, holistic sense, where complete languages are characterized as 'isolating', 'fusional' or 'agglutinating' – there is no empirical basis for this typology; see e.g. Bickel & Nichols 2005, Haspelmath 2009): there is voice affixation (which plays an important role in lexical classification in Tagalog – see Himmelmann 2007).

- p.33, example (19), p.43 ex. (43) (among others): I understand that it is not regarded as a crucial issue in the advocated approach, but it may be worth noting that there are cases in which an action-denoting lexeme used in a referring function does not shift to person- or thing-denoting semantics at all, but rather still refers to the action 'as a whole'. See also special issue of Theoretical Linguistics 38: 1/2 (2012).

- p.34/35, criterion 2/2': I don't understand how 2' can be a rephrasing of 2. Just because a particular category is marked synthetically or analytically doesn't entail that it is determined lexically or syntactically, respectively? By the way, it seems that two aspects of morphological typology are conflated here: the degree of phonological connection between base and affix (fusion) and the amount of material packed into a single word (synthesis).

- p.55, final line of paragraph starting with *Evans and Osada*: For info: one of the papers in Rijkhoff & Van Lier (fc 2013) by David Gil does this for Riau Indonesian.

TYPOS

- p.2, penultimate line: *ad nauseam*
- p.18, 3rd line above paragraph starting with *Analyzing*: delete *it*
- p.19, 2nd line of final paragraph: add space before *but*
- p.31, final word above section 4.1: add section number
- p.34, 3rd line of 4.1.2: *than* → *then*

REVIEW 2

The revised version is considerably improved compared to the first version. The ideas are explained in greater detail. Competing theories (Hengeveld, Croft, Evans & Osada) are discussed in a little more detail, though they are not properly presented to the reader as alternatives. The resulting paper is hence much better written. It is also 70 pages single-spaced in draft form; it would probably be longer in published form.

The most important question, then, is the content. Is the content sufficiently interesting to merit making the MS one of the longest articles to be published by LT? Here I am inclined to say "no". This is a somewhat tricky recommendation; most MSS that I recommend rejection for are poorly written as well as having significant problems in terms of content, or represent superficial treatments of the phenomenon in question. I have two reasons for remaining dubious about publishing the MS, one having to do with what the paper claims to achieve and one having to do with what it actually achieves.

The explication in the revised version leads me to think that the MS presents a canonical typology of lexemes masquerading as a canonical typology of syntactic categories or parts of speech. Although the author compares his/her approach to that of Hengeveld and Croft, and discusses the "big three" functions (reference, predication, modification), s/he also points out that the canonical typology s/he presents applies to constructional distributions of all sorts (including such "constructions" as inflectional and derivational affixation - which are indeed constructions in construction grammar). I think that this is an accurate representation of the author's canonical typology. The following statement on p. 21 indicates that this is the right interpretation: "Many of the criteria for determining lexical flexibility are in fact criteria for assessing the extent to which different words count as the same lexeme". But it does mean that his/her goals are quite different from those of Croft and even Hengeveld. (They are perhaps closer to those of Evans and Osada.)

Croft's theory of parts of speech is about the relationship between what he calls the propositional act functions (reference, predication and modification) and

semantic classes of words. Hengeveld's focus is similar to Croft's, in that he is interested in how words fit into these three functions, and is explicitly a theory of categories based on those functions - Verb, Noun, Adjective, Adverb. Both Croft and Hengeveld are interested in the constructions that express reference, predication and modification, and how words are distributed in these constructions, with the aim of elucidating the theoretical concepts of "noun", "verb" and "adjective", even though their theories are quite different. The author's theory, however, is explicitly about lexemes, based on a specific assumption about lexemes, namely, that all semantically related uses of a word form constitute a lexeme. The author states that most of the exemplification of the canonical typology are with the "big three" functions, but that it applies in principle to any type of construction. At least this is what the author says; I do not know how far s/he will follow this through. For example, English has many words like "chocolate" which can occur in an uncountable construction ("a piece of chocolate") and a countable construction ("two chocolates"), with a variety of semantic relationships between the two uses. So English would be example of a fairly "flexible" language in this regard, even if it less flexible with respect to the major parts of speech.

This difference between the author's theory and the supposedly competing theories is not very clear in the MS, because the competing theories are not actually presented, so that the reader who is not already familiar with them would have no idea of what they are about, let alone how they differ from the author's. Instead, the author picks out certain assertions by the other scholars and either agrees or disagrees with them. In fact, it is only with the discussion of the author's view on lexemes and semantics in the revised version that I inferred that the author is really doing a canonical typology of lexemes. This is orthogonal at least to Croft's theory of parts of speech, since Croft makes no reference to lexemes, just to semantic classes of words.

The author takes a strong - some would say, extreme - position on lexemes: any semantically related use of a root or word form is part of the same lexeme. Hence the author cannot really say s/he is taking a "neutral" position. On the other hand, if one were to do a canonical typology of lexemes, then one would indeed cast the net broadly, as the author has done; just as Corbett casts the net very broadly for "agreement" in his canonical typology of agreement. This is why I suggest that the author is presenting a canonical typology of word formation, or perhaps of derivation (see for example the discussion at the bottom of p. 55): what are the different ways that different uses of a word form or root are expressed in languages? - without committing to a particular theoretical position on lexemes.

This leads me to the second issue I must raise with respect to this MS. If we (that is, the author, the editor and at least this reviewer) agree that this MS is really a canonical typology of lexemes (or whatever), and therefore has different theoretical aims than theories of parts of speech like Croft's or Hengeveld's, how should it be evaluated as such? And here I have serious reservations, particularly for such a long MS. A canonical typology is basically a set of parameters that can be used for a typological classification of some linguistic phenomenon, like agreement or lexemes. It is claimed to differ from run-of-the-mill typological classification found in typology textbooks or typological research monographs in

that it is supposed to be a "neutral" set of parameters organized in terms of a focal type on all the parameters, namely the canonical type. In practice, it does not seem that different from good examples of "run-of-the-mill" typological classification; in fact, canonical typology does not seem to grapple with the problem of defining formal typological traits crosslinguistically, which has been discussed in the typological literature from Greenberg 1966 onwards.

But that is not the main issue. In the end, a canonical typological classification, even if it is superior to a "run-of-the-mill" typological classification, is still just a classification of crosslinguistic data. What we are ultimately interested in are typological generalizations, that is, universals of language, and plausible explanations for those universals. Even Corbett's canonical typology of agreement has the Agreement Hierarchy, a robustly attested language universal, to draw our attention (a universal that was incidentally discovered by Corbett long before he developed canonical typology). Croft and Hengeveld both propose implicational universals for parts of speech as part of their theories. Unfortunately I don't see any such universals in the MS. Would LT publish a 70+ page typological classification of some grammatical phenomenon, without any generalizations observed and explanations proposed for them? (An additional consideration in this regard would be if the MS pulled together a wide range of types, including types not yet discussed in the typological literature on the topic. But the examples are mostly existing examples, or examples similar in type to those already discussed by Hengeveld, Croft and Evans & Osada.)

The lack of generalizations also entails that there is a lack of motivation for the parameters chosen by the author for the canonical typology in the MS. Why the set of features chosen, and why is a particular configuration of feature values the "canonical" one? The features chosen are not really "neutral" or "aprioristic" (p. 13). The parameters are partly empirically induced - what properties of the phenomenon are observed to vary across languages? (This is no different from "run-of-the-mill" typological classification.) I am also not sure that all linguists would agree on the "neutral" set of parameters. The values of the parameters that constitute the canonical type (or the canonical opposing types, see Fig. 2 on p. 9), represent some empirical convergence of parameters, even if the convergence of all values is nonexistent (often for good communicative reasons, as the author notes late in the MS).

Also, many of the parameters are not fully logically independent of each other. Here is a summary of the ten parameters:

1. Indistinguishability (structural coding of different functions)
2. Indistinguishability of inflection
3. Indistinguishability of distributional potential - but part of distributional potential is occurrence with different inflections, i.e. this parameter subsumes #2
4. Indistinguishability of functions - it's not clear to me how this differs from #1; both have to do with occurrence of lexemes in different propositional act functions/constructions
5. Indistinguishability of feature values - this is unclear and seems highly specialized; the only examples contrast inherent features (of nouns) vs. index features (indexation associated with predicates and modifiers)

6. Indistinguishability of meaning across functions - this is difficult to measure, to say the least
7. Indistinguishability of semantic shift across lexemes - unclear how this differs from #6
8. Scope of flexibility across functions - this appears to be subsumed by #4
9. Scope across lexemes - sort of a meta-parameter of flexibility summarizing over #1-#8.
10. Scope across lexemes and functions - just a combination of #8 and #9. It looks like an attempt to fill the list out to exactly ten items.

Finally, as discussed in some of the specific comments below, the view of semantics and the semantics-pragmatics debate seems a bit old-fashioned, that is, it does not take into consideration more recent theories such as the usage-based model. Likewise, the comparison of canonical typology to prototype theory is inaccurate; canonical typology is a prototype theory, but it is not an exemplar theory (see below).

Specific comments

p. 7, sec. 2.2: English nouns do not "resist use in predication". Evans and Osada do a quantitative comparison of English and Mundari, and the two languages do not come out that differently. I think this is just the bias that we are all "certain" that English has sharply distinguished major word classes (N, V, A). What is interesting is how nouns can be used in predication. Clark & Clark (1979, *Language*) remains an insightful study here.

p. 7: why is lack of plural marking an indication of noncanonicity?

p. 10: the IPA vowels are not the best choice. What has in fact happened is that language vowels are mapped onto the F1-F2 acoustic space by those who are seriously interested in comparability of vowels across languages.

p. 11: the transitivity parameters covary because the features used are not logically independent. For example, punctual entails telic, and telic entails process.

p. 12, Table 1: it is not clear to me how this "canonical typology" of possession is different from "ordinary" typological classification.

p. 13: I don't know what a canonical typology within a single language would be.

p. 14: "whether they [parameters] are formal or functional is of little relevance" - but for many typologists, many if not all formal categories are language-specific. There is a big different between formal and functional categories and they should not be confused with each other.

p. 14: not all theories of categories treat prototypes as existing exemplars. In fact, psychologists standardly distinguish between prototype theories and exemplar theories. Canonical typology is actually more like a prototype theory than an

exemplar theory. See Murphy, "The big book of concepts" (MIT Press, 2002).

p. 15, fn6: all these properties are language-specific, so it's hard to see how they could be part of a typological theory of parts of speech (this is perhaps more a criticism of Sasse 2001 than of the author).


p. 16, fn 7: this is somewhat of a misconstrual of Dryer 1997. He talks about language-specific grammatical relations, but also of crosslinguistic variation in those grammatical relations, and functional-cognitive explanations for that variation.

p. 17: both Croft's and Haspelmath's are functionally based (reference, modification/narrowing reference), so these are not really about language-specific constructionally-defined word classes. Haspelmath's quote is explicit about this.

p. 17-18, sec. 3.2: this is the critical issue. The author takes a lexeme to be all word forms with related meanings. In this respect the author's approach is completely different from Croft's (and not necessarily incompatible). Croft is only interested in semantic classes of words; if a word has another meaning, then that meaning belongs to another semantic class, and so Croft's generalizations do not apply. Croft doesn't talk about lexemes. And this is the real point: the author's theory is a theory about lexemes, broadly construed as the author does to include all semantically related uses of the word form (systematic or idiosyncratic). So it's about a different linguistic phenomenon than Croft's or Hengeveld's theories.


p. 23: Pace the author, Croft argues against both lumping and splitting approaches to syntactic categories.

p. 27: it is not clear to me that having a smaller number of color terms to cover the same color spectrum is "lexical flexibility" in the same way that a language that allows a word to express a property, a thing possessing the property, and coming to have the property is said to be flexible.

p. 28, line 4, "If all ": this is unclear; do you mean distributional criteria within the language?

p. 28, definition of flexible category: this definition is so broad that any language and probably any lexeme is flexible. If functions are defined in terms of the constructions in which a lexeme occurs, then lexemes are flexible by this definition if they occur in more than one construction - which all lexemes do. See also the discussion at the bottom of p.33, on the flexibility of Inupiaq lexemes.

p. 29, "no distinction in form, meaning, or syntactic possibilities" - what is meant by "syntactic possibilities"?

p.45, "lexemes like -ganga 'doctor/cure'  could mean anything having to do with curing or cures" Really? Can -ganga mean 'aspirin'? 'druggist'? 'hospital'? 'Medicare'? 'operating room'? Where do you draw the line for "having to do with..."? The discussion on p. 46, paragraph 2, is just hand-waving. More generally, this is a very old-fashioned view of meaning and context.

p. 48, ex. (40): this "formula" leaves out the most problematic case, namely conventional semantic shift (particularly idiosyncratic shift) in zero-conversion.

p. 48, "conventionally predictable": this is a contradiction in terms, since convention is by definition at least partly arbitrary. (The same confusion is found on p. 50: "a matter of linguistic convention, and so in a certain sense are just as predictable...") This paragraph generally misunderstands the debate, namely conventionality vs. pragmatic predictability.

p. 49, ex. (41): the semantic shifts are described as "completely predictable". If so, why doesn't laTab mean 'forge into scissors' and kaTu mean 'cut with a knife'?

p. 50, Table 6: this table is titled "idiosyncratic semantic shift for derivation in Fijian", but it appears to be about English, not Fijian

p. 50: canonical typology is not neutral.

p. 63, fn 14: this footnote sweeps under the rug the fact that certain Lango PC lexemes DO differ grammatically from other PC lexemes - and they are concepts in Dixon's property concept prototype. Yet the canonical approach says nothing about why this should be the case.

REVIEW 3

First time around, I recommended publication subject to revision, and my main concerns at that stage – revolving around the notion of lexeme – have now been met with a much more worked-out discussion. So I think the paper should definitely be published, but there are still some revisions that are needed.

Perhaps the most important concerns size: it has blown out from twenty-something to seventy pages. This is way too long for a paper of this scope and will diminish its readership, so it needs to be pruned right back. There are obvious places where this can be done: the discussion of canonical typology runs to ten pages and could be compressed to a page or two since discussions of this approach are now widely available (some bits, such p. 11-12, are easy to sacrifice; achieving the compression of the rest of this section will obviously be trickier). There is also a great deal of unnecessary repetition, such as the recurrent use of the term 'methodological opportunism', as well as discussions that wander far from the main point, such as the discussion of mother-in-law speech in the first half of p. 52. Taking a ruthless knife to slim down the prose will make for a much tauter and more readable piece.

There is still some unclarity in the discussion of the lexeme. For example, on p. 18 it is said that calling Mundari buru n. 'mountain' and buru- v.t. 'heap up' 'separate lexemes' 'ignores the cognitive connection between them'. In fact there is very little consistency among lexicographers in the use of the terms lexeme and e.g. lemma, but there are many traditions which explicitly allow the linking of lexemes by various 'semantic bridges', thus removing the impediment to discussing a cognitive connection.

Given that one of the strengths of the paper is the more interesting way it offers for dealing with cross-linguistic variation in how much conversion is found across word-classes (whether by zero-derivation or overt means) it is a shame that there is not more discussion of how to quantify this. The paper gives a large number of dimensions on which languages can be flexible, but virtually no indication of how we might measure this (except for a few places, such as the discussion of percentages of lexemes that can convert).

There are also a number of minor errors and carelessnesses which need rectification. Murrinh-patha is misspelt as Mirrinh-Patha (p 15) and disappointingly the primary source for the data and discussion (Walsh 1996) - an early discussion of part-of-speech flexibility or at least gradience - is omitted in favour of a secondary one (Sasse 2001). More carelessly, at one point (bottom of p. 49) a discussion by Dixon about English is represented as a discussion about Fijian (to make matters worse, what should be Dixon 2012b:47 is mis-cited as Dixon 2012a:48)! On p. 31 it is stated the 'the means of distinguish[ing] nouns and verbs] in Mundari are extremely subtle' - this is simply inaccurate, given the rich verb morphology in Mundari which cannot apply to nouns. There are other odd statements, such as the one on p 32 that Tagalog is 'isolating', clearly false in the light of its intricate verbal morphology.

More specific comments

p. 4, para 1. 'where no such distinction'... Well, the meaning difference is already there quite clearly.

p. 35, ex. (22). Turkish example needs more discussion: to say 'this is the case for the majority of Turkish lexemes' left me wondering what is the case - allowing abstract quality nouns, adjectival attribution, and manner adverbial use?

p. 51, mid 'as for Spanish' - pres. -migo and -tigo

p. 54 Be consistent in when to hyphenate

p. 56. I find 'monocategorial' better than 'monocategorical'

Many minor misspellings and unclarities, e.g. Saddock (> Sadock)

Frans Plank
Sprachwissenschaft
Universität Konstanz
78457 Konstanz
Germany

Tel [+49 \(0\)7531 88 2656](tel:+4907531882656)
Fax [+49 \(0\)7531 88 4190](tel:+4907531884190)
eMail frans.plank@uni-konstanz.de
<http://ling.uni-konstanz.de/pages/home/plank/>

--

Omnis habet sua dona dies.
~ Martial