

Causation as a Factor and Goal in Typological Comparisons^{*}

Randy J. LaPolla

Nanyang Technological University

Background:

For some time in the field of typology there has been debate about the nature of the categories or concepts used in typological comparison and the difference between these categories or concepts and descriptive categories (the categories used in describing individual languages).

In early 2016 an extensive debate on this topic on LingTyp, the listserv list for the Association for Linguistic Typology,¹ resulted in a special issue of the journal *Linguistic Typology* (20.2) on this topic with position papers written by some of the members who had participated in the online debate, in response to a series of questions from the Editor of *Linguistic Typology*. There was no clear resolution to the differences between the two main camps involved in the debate, but a follow-up to that debate earlier this year (2017) led to pinpointing the key issue that divides the two camps, causality, and I'd like to present it here as an important issue we should discuss.

Summary of the January 2016 debate:

Through much discussion of methodology in descriptive and comparative studies, it became clear that we all agree that language description should be inductive and based on the facts of the language, that there are no cross-linguistic formal categories (see e.g., Dryer 1997, 2016; Croft 2001; Haspelmath 2007, 2010, 2017), that language-specific descriptive categories used in describing a language are unique to that language, and that comparative categories are

^{*} I'd like to thank Martin Haspelmath, Frans Plank, and two anonymous members of the Editorial Board of *Linguistic Typology* for helpful comments on this note.

¹ The full discussion is available at

<https://www.degruyter.com/view/j/lity.2016.20.issue-2/lingty-2016-0007/suppl/lingty-2016-0007ad.zip>

and in the LingTyp archives at

<http://listserv.linguistlist.org/pipermail/lingtyp/2016-January/thread.html>

under the heading “Structural congruence as a dimension of language complexity/simplicity”.

abstractions away from the actual details of the language-specific categories, and it seems the major issues on which we differ are (a) how far the abstractions can go from the actual facts of the language, and (b) exactly what are “the facts of the language”. The reason for the latter issue is that Matthew Dryer and Martin Haspelmath argue that there is a difference between the grammar of a language and the usage manifested in the texts of the language, essentially the Structuralist distinction between langue/competence vs. parole/performance, and argue that typology is based on the latter, and not on the former, which means looking at language structures without reference to the conventions and motivations that produced them.² Jan Rijkhoff, Peter Arkadiev, William Croft, and I independently argued in the debate that the usage *is* the facts of the language, and *is* the grammar of the language, and so there is no disconnect, and so there is no justification for ignoring the reasons for the patterns being the way they are.

At the time, I assumed the major difference between the two camps was just in how far we are willing to go in our abstractions, and so in my paper for the *Linguistic Typology* special issue on this question (LaPolla 2016), I argued that in doing both description and comparison we should work inductively, staying true to the facts of the languages as manifested in natural data, and not resort to extreme abstractions that lead to classifying languages or constructions in a way that obscures the diversity of the structures of the languages.

The March 2017 debate:

In March this year the issue came up again on the LingTyp list, as Martin Haspelmath proposed to group gender systems and classifier systems into one comparative category, called “genifiers”, and again some of us opposed categorisation that we argued obscured the diversity of the structures, as we feel gender marking and classifiers are doing different things.³ Haspelmath and I later had an extended discussion of this off the list, and he sent me a paper he had just finished, a draft of his reply to the papers in the *Linguistic Typology* special issue on classification. In that paper, he points out how his use of “the facts of the language” and my use of “the facts of the language” (as in the title of my 2016 paper: “On categorisation: Stick to the facts of the language”) differ. He says that what I am talking

² Although Haspelmath and Dryer’s view assumes the same competence/performance distinction as that of the Chomskyan Structuralists, it is the flip-side in terms of which is considered important, as Haspelmath and Dryer argue that performance is what we are interested in, unlike Chomsky, for whom only competence is of interest.

³ It is also a general principle that the more abstract the category the more there is a loss of information about the diversity of structures.

about is not the facts of the language, but the rules of the language, whereas he is talking about the actual forms as spoken without regard to why they are the way they are, and argues that is what typology is interested in. This again reflects his Structuralist assumption of a distinction between langue/competence vs. parole/performance, as he mentions in a footnote (Haspelmath 2017, §10, p. 19):

“... the universal in (10a) entails a statement such as (10b).

- (10) a. In almost all languages, the subject normally precedes the object when both are nominals. (Greenberg 1963, Universal 1)
- b. In Mandarin Chinese, the subject normally precedes the object.

LaPolla (2016: §2) objects to the claim that Chinese is an SVO language (which is a more specific claim than (10b), but otherwise very similar) because he has shown in earlier work that Chinese does not have any subject or object category, and he thinks that “labeling [Chinese as an SVO language] implies that these categories either determine word order or are determined by it” (cf. LaPolla & Poa 2006). But again, this is not so. (10b) is a correct factual statement about Mandarin Chinese (assuming that “subject” means S/A, and “object” means P), and it is not a rule of Mandarin grammar.* LaPolla may be right that “most people who see a description of Chinese as SVO will in fact assume that the label was given to the language because those categories are significant for determining word order in the language” (2016: 370). But if they do, they have not understood the difference between describing a language and classifying a language from a comparative perspective. These two are different enterprises – not completely unrelated, because both are based on the facts of the language, but also not identical . . . Thus, what we compare across languages is not the grammars (which are incommensurable), but the languages at the level at which we encounter them, namely in the way speakers use them.

*Footnote: Confusingly, LaPolla (2016) uses the expression „the facts of the language“ in the sense in which I use „rules of the language“ (this strange terminology may be motivated by his rejection of „structuralism“ and the competence/performance distinction).”

In the discussion we had that followed, the issue came down to whether we think one needs to look for causality or not. My view is that, for example, making the statement of (10b) in terms of A/S or “subject” makes it seem as if that is the causal factor, i.e. the word order is the way it is because it is determined by an A/S pivot or “subject”, so I would prefer

statements that really do include the clausal factor, e.g. information structure in the case of Chinese. If a language has grammaticalised A/S pivots in a majority of constructions in the language, then one might be able to make a statement that the word order reflects this A/S pivot, leading to the word order pattern we find most often, but if there is no evidence of that, then other factors should be identified as resulting in the patterns found. The facts that explain the patterns found in the languages then should be operative in the comparisons, rather than the comparisons being done simply on a string of words with no reference to why the words are in the order they are. This is not arguing for comparing descriptive categories, but taking into account the causes of the patterns manifested.

Haspelmath, on the other hand, said (email, 29 April): “I never use the term ‘causal factor’ at the level of language-particular conventions. Speakers obey rules because they are conventional, not because they are functional. I think functional considerations (and ‘causal factors’) come in only at the level of cross-linguistic generalizations . . .”⁴

And again, (8 May): “. . . I think the comparisons should be neutral with respect to possible explanations – they should involve only the conventions, or the behaviour of the speakers, but not the explanations, because I think that it's too hard to talk about explanations at the particular level.”

And on 14 August, in reply to a draft of this paper: “. . . there are different kinds of causal chains . . . What I do in my work is to restrict the term "causal" to "causal factors" that explain universals, as in ‘higher frequency of singulars -> greater length of plural forms’ (it is a very strong universal trend that plural forms are longer than singular forms) . . .”

The term “causation” may not be the best for what I mean, as it can lead people to relate it to things that I don’t intend, as Haspelmath initially did. What I mean by causation is why the speaker’s output is the way it is, what caused it to be that way, rather than just looking at the output without reference to why it is the way it is, as Haspelmath and Dryer argue. This

⁴ One of the Editorial Board members who commented on a draft of this paper pointed out that Haspelmath’s view here is in conflict with a foundational assumption of psycholinguistic research on language production, “that speakers continuously and implicitly choose between different word orders or structures (active/passive, ditransitive alternation, etc.) based on processing-related (i.e., causal) factors . . . Similarly, psycholinguistic research points to phonetic and phonological realizations to be affected by causal factors on a moment-by-moment basis”. I can add that it is also in conflict with the assumptions of Interactional Linguistics, where the linguistic structures are seen as emerging from the joint process of the interactants trying to achieve their goals in their interactions (see, for example, Hopper 2011).

would hold for many things, non-linguistic as well. In phonetics, for example we can look at a sound wave produced by a speaker, and compare sound waves across languages, for example saying that this language and that language both have lowered 3rd formants on some vowels, but that doesn't tell us why the formant is lower, just that it is. There are several reasons why the 3rd formant might be lower, so we would want to see what the cause of the lowering was and then compare the languages in terms of which ones have lowered 3rd formant due to tongue retroflexion, which ones have it due to tongue root retraction, which ones have it due to lip protrusion, which ones have it for some other reason, and see what features are concomitant with that particular articulation. A non-linguistic example would be if two patients presented at a doctor's clinic with a cough, we would want to first determine why they were coughing, and possibly find that one had black lung while the other simply had a cold, rather than just say they both have a cough and so should be treated the same way.

Summary of the differences in our positions:

Martin Haspelmath and Matthew Dryer argue that

- a) in doing comparisons one can in most cases ignore the morphosyntactic facts of the patterns found (i.e. do comparisons on a largely semantic basis, as Matthew argued that the morphosyntactic nature of the forms does not affect the relevant patterns—see Dryer 1989),⁵ and
- b) one can ignore the causes of the patterns being compared, and just directly compare the forms without consideration of why the forms are the way they are.

I argue that

- a) it is necessary to take into account the actual morphosyntactic realisation in, for example, constituent order studies, because in the languages I work with the nature of the morphosyntactic element does influence the pattern it manifests, and

⁵ Although in Dryer 1992: 120 it is explicitly stated that how semantic categories are realised in particular grammars is relevant to the patterns of correlation, in his post to the LingTyp list on 19 Jan 2016 as part of the debate on comparative categories, Dryer says. "I have spent considerable time over the years collecting data, not only on word order in particular languages, but also on the syntactic realization in these languages, precisely to examine empirically whether the syntactic realization makes a difference. The result is that while the syntactic realization sometimes makes a small difference, it is overall irrelevant: by and large, generalizations over semantic categories apply the same, regardless of the syntactic realization."

b) we need to consider the reasons why the pattern is the way it is, otherwise you end up comparing very different systems that just happen to produce similar-looking results. (E.g. a rocket's propulsion system and the wings of a bird both result in the object flying, but classifying the rocket as a winged entity because the result (flying) is the same as for entities with wings does not make sense, as the reason for the flying is completely different.)

To go back to our original example, word order in Mandarin Chinese is determined by information structure, not by grammatical relations or mood, while in English word order is determined by grammatical relations and the marking of mood.⁶ From Matthew Dryer and Martin Haspelmath's point of view the two languages can be lumped together in a category called "SVO" (e.g. Dryer 2013) just because they happen to both end up with a verb-medial pattern in many clauses. That is, having seen that English is verb medial because it is SVO, they then say that any language where the speakers output a verb medial pattern can then also be categorised as SVO. As with the rocket example, I don't think that this is a useful way to do comparison. Ignoring the causes can lead to problems like the rocket being classified as a winged entity because it flies, like entities that actually have wings. I also think it obscures the diversity of linguistic forms, and of course doesn't address causality at all. This relates to our goals in doing comparative linguistics, and it may be we have different goals. For me the goal is to investigate the diversity of patterns in languages and determine the reasons for the patterns being the way they are, and so the more finely detailed we can get in understanding the diversity the better, and that is why I oppose lumping languages together in a way that obscures the diversity of structures and their motivations.

To use Martin's example of 'higher frequency of singulars -> greater length of plural forms' as a causal chain,⁷ in order to be able to do the comparisons that would allow us to test this, we would first have to identify which forms are plurals in the languages; we cannot say that English has -s for plural, and so if we find -s at the end of nouns in some other language it can be classified together with English plurals because the output (Haspelmath's "facts of the language") is the same. That, is, we need to take into account the language specific categories of the language. We need to do that also to determine in a particular language which

⁶ See LaPolla 2009 on Chinese clause structure, and Michael Halliday's brilliant analysis (e.g. 1994) of why English structures are the way they are. See also LaPolla & Poa 2006 for a summary and comparison of the factors that determine clause structure in English, Chinese, and Tagalog.

⁷ I actually would not see this as causation, just correlation which requires an explanation in terms of causation.

functions a form has when the form has several functions. For example, in Tagalog, reduplication of the initial syllable of a root in some cases marks plurality, but in others marks imperfective or what has been called “contemplative”. So we cannot take the form at face value (just look at the output), we have to know why it is the way it is, what is its function, in order to do the comparisons.

When I finally came to understand Haspelmath and Dryer’s position this past May, it came as quite a revelation and a shock, as all along I had assumed all linguists were interested in causation, that is, why the forms in languages are the way they are.⁸ In a 2002 paper criticising methodology and explanation in word order studies, I criticised the type of implicational universal given in Martin’s (10a) above for (among other things) showing only correlation, not causation. But I was missing the point that they were not trying to show causation, but only correlation based on superficial formal or simply semantic resemblance. This issue is not just one of methodology, but relates to our most fundamental notions of what language is and how communication works, and what the goals of comparative linguistics are. This note is to encourage discussion of this issue and the related underlying issues.

References

- Croft, William. 2001. *Radical Construction Grammar: Syntactic theory in typological perspective*. Oxford: Oxford University Press.
- Croft, William. 2003. *Typology and universals*. 2nd edn. Cambridge: Cambridge University Press.
- Dryer, Matthew S. 1989. Discourse-governed word order and word order typology. *Belgian Journal of Linguistics* 4: 69-90.
- Dryer, Matthew S. 1992. The Greenbergian word order correlations. *Language* 68.1: 81-138.
- Dryer, Matthew S. 1997. Are grammatical relations universal? In Joan L. Bybee, John Haiman & Sandra A. Thompson (eds.), *Essays on language function and language type*, 115–143. Amsterdam: Benjamins.

⁸ Dryer, in his excellent contribution to the LT special issue (2016: 315), says “The differences among languages arise because of competing motivations (Haiman 1983; Du Bois 1985; Croft 2003), where different explanatory forces compete with each other and languages differ because different explanatory forces win out in different languages.” Yet in doing comparative work he argues for ignoring these explanatory forces.

- Dryer Matthew S. 2013. Order of subject, object and verb. In Matthew S. Dryer & Martin Haspelmath (eds.), *The world atlas of language structures online*. Leipzig: Max-Planck-Institut für evolutionäre Anthropologie. <http://wals.info/chapter/81> (accessed on 13 March 2016).
- Dryer Matthew S. 2016. Crosslinguistic categories, comparative concepts, and the Walman diminutive. *Linguistic Typology* 20.2: 305–331.
- Du Bois, John A. 1985. Competing motivations. In John Haiman (ed.), *Iconicity in syntax*, 343–366. Amsterdam: Benjamins.
- Haiman, John. 1983. Iconic and economic motivation. *Language* 59: 781–819.
- Halliday, M.A.K. 1994. *An Introduction to Functional Grammar, 2nd edition*. London: Edward Arnold.
- Haspelmath, Martin. 2007. Pre-established categories don't exist: Consequences for language description and typology. *Linguistic Typology* 11: 119–132.
- Haspelmath, Martin. 2010. Framework-free grammatical theory. In Bernd Heine & Heiko Narrog (eds.), *The Oxford handbook of grammatical analysis*, 341–365. Oxford: Oxford University Press.
- Haspelmath, Martin. 2017. How comparative concepts and descriptive linguistic categories are different. ms, March 2017.
- Hopper, Paul. 2011. Emergent grammar and temporality in interactional linguistics. In Peter Auer & Stephan Pfänder (eds.), *Constructions: Emerging and emergent*, 22-44. Berlin: Walter de Gruyter.
- LaPolla, Randy J. 2002. Problems of methodology and explanation in word order universals research. In Pan Wuyun (ed.), *Dōngfāng yǔyán yu wénhuà [Eastern language and culture]*, 204–237. Shanghai: Dōngfāng Chūbǎn Zhōngxīn.
[LaPolla 2002 Problems of Methodology and Explanation in Word Order Universals Research.pdf](#)
- LaPolla, Randy J. 2009. Chinese as a Topic-Comment (not Topic-Prominent and not SVO) language. In Janet Xing (ed.), *Studies of Chinese linguistics: Functional approaches*, 9-22. Hong Kong: Hong Kong University Press.
- LaPolla, Randy J. 2016. On categorization: Stick to the facts of the languages. *Linguistic Typology* 20.2: 365–375.
- LaPolla, Randy J. & Dory Poa. 2006. On describing word order. In Felix Ameka, Alan Dench, & Nicholas Evans (eds.), *Catching language: The standing challenge of grammar writing*, 269-295. Berlin: Mouton de Gruyter.