Report on the PhD thesis *"Teoretické nástroje pro kategorizaci a funkční explanace vztahů v jazyce"* (English translation: *"Theoretical tools for categorization and functional explanations of relations in language"*) by Mgr. Jan Křivan, Charles University

Summary

The main aim of this PhD thesis was to explore a fundamental question in linguistics: Why are languages the way they are? How can we best explain the shape of linguistic structures and what is the best approach to the study of linguistic categories? The author approaches this question from a functional-evolutionary perspective, and seeks to show that categories emerge in language in response to communicative constraints associated with use. He empirically tests this hypothesis on the category of possessives and the comparative in large corpora of spoken and oral Czech. The results, the author argues, show that the structures are functionally driven and most compatible with the principles of the usage-based approach (e.g., Bybee, 2007), providing evidence for language as a phenomenon of third kind (Keller, 1994).

I recommend this thesis for an oral defense.

Evaluation

I really enjoyed reading this PhD thesis. It is well written and well structured, and it has deepened my interest in some of the fundamental assumptions and questions in general linguistics, but particularly those associated with the functional-evolutionary approach promoted in this work.

The theoretical background and motivations for the study, the questions and methods, are thoroughly presented in Chapter 2. Chapters 3, 4 and 5 provide the main content of the thesis with novel empirical data and analyses. The results of the study on possessives in Chapter 4 were particularly insightful and innovative, and I applaud to the thorough and multifaceted corpus approach to the distribution of different possessors. I agree with the candidate's overall interpretation of the data and the evidence for the role of information structure in Czech possessives. I see a great potential in his corpus approach (as well as in

his intuitions) but find it unavoidable to address some of the remaining questions with experimental studies (as the candidate himself acknowledges, p.148). One obvious issue to examine would be the identity of the MLP anchor (i.e., speaker/listener/or grammatical subject, as suggested by the author?) in both production and comprehension experiments.

What I found a bit disappointing, however, was the unequivocal interpretation of the prenominal possessor data in favour of the animacy hierarchy. I believe there is space for further discussion. One aspect that stands out and requires some explanation is, for example, the total lack of neuter nouns referring to humans. The lack of neuter nouns is acknowledged at the onset of the chapter (p.76); this is stated as a fact, a result of 'morphosyntactic constraint'. But what is the source of this constraint and how can we account for this 'anomaly' in the animacy hierarchy? Is this lack of neuter nouns with human referents driven by, for example, low frequency of type (or token) or by some internal analogy within the morphological system since there are no neuter proper names? Any of the above interpretations would be compatible with some functional explanation but not simply by 'animacy'. Furthermore, it would seem appropriate to show how pronouns fit the animacy hierarchy since, clearly, they can refer to non-human animate entities but also inanimate entities? Wouldn't the modified accessibility hierarchy of 'expression type' (i.e., O'Connor et al., 2013) describe this aspect of the data better? It is of course possible that animacy is in this work conceptualized differently, but if this is the case, then it should be better communicated.

The findings in Chapter 5 highlight variation of morphological suffix in the Czech comparative as a result of frequency. The goals, methods and results were clear, although I did not find it all convincing, primarily because there were no details on some of the operational definitions and coding (e.g., 'reduction'). I would have also preferred a better, more concrete, explanation of why the results provide evidence for Keller's model of language as a phenomenon of third kind. I would therefore suggest that the candidate reviews some of the critical concepts from earlier parts of the thesis, particularly from Chapter 2, to refresh the reader's memory in order to substantiate his claims and interpretations. I would further suggest that he illustrates his hypotheses with concrete examples

(see e.g., Section 5.3, p. 174). Finally, in terms of interpretations, the chapter was strictly focused on the contribution of frequency but some findings indicate a potential role of other factors in the patterns of variation (see e.g., phonological analogy in adjectives like 'hezký' *vs* 'pěkný' that patterns with 'krásný' and 'nadherný'). I wonder what predictions can be made for its interaction with frequency. With the data from this and previous chapter, the candidate can also return to the discussion of the status and role of exceptions in language categorization as raised in Chapter 2.

General observations

There were a few areas of the thesis that would, in my opinion, benefit from further development. The thesis does not clearly articulate what is novel about this work, and how it advances our knowledge and understanding of language and linguistics. It is critical that the author clearly distinguishes between previously established views and newly developed ideas. The novelty, as it is now, appears limited to the empirical data more than to their interpretation and their impact on theories of language structure. This lack of explicitness in communicating novelty is, I reckon, possibly a result of different cultural styles and conventions, but it should be fixed.

A further aspect of this thesis that can be improved is the motivation behind some fundamental assumptions and decisions. Methodologically, it was not explained, for example, what motivated the choice of the two particular structures examined in this thesis: why possessives and why the comparative? Why are these structures suitable candidates for this type of investigation? Why not others? The thesis would become stronger (and the studies more coherent) if the choice of these structures is clarified.

Conceptually, the work would be more accessible for the reader and, at times, more convincing if the author re-iterated his assumptions and arguments with concrete examples in each empirical chapter and in the final chapter. In general, my feeling was that the author relies on the reader's knowledge of the topic and the field a bit too heavily. The content of the final chapter would become stronger if there was a more thorough integration of the novel findings with previous work, particularly as related to the content of Chapter 2 – in light

of theories that are in favour but also against the usage-based view of language adopted here.

General questions for discussion

I also have a few questions that are related to this research and the views expressed in the thesis. One question I have already hinted on earlier in my report is the motivation behind the selection of the two categories: possessives, on the one hand, and adjectives or rather the comparative, on the other. The two categories are distinct. One of the distinctions is that it is easy to define one category, i.e., 'the comparative', but, as the author acknowledges, it is much more challenging to agree on what a possessive is. Why is the category of possession so different from the comparative? What is the relationship between 'form and function' or morphology and semantics or, even more broadly, language and cognition? And what role, if any, does this distinction play in this research and the results?

The focus of this PhD thesis is related to the origin and nature of categorization in language. The thesis exemplifies the role of language use on two specific constructions. How generalizable are these results? Do you expect the same forces visible across different language domains, phenomena, and units of structure? Why?

The candidate adopts a usage-based approach for the study of language structure and change, advocating a systematic/consistent approach to data and categorization. But my feeling, while reading the thesis or elsewhere, is that there is, in even in the context of the usage-based approach, space for speculation and *post-hoc* analyses. What is the reason that the usage-based and typological-evolutionary approach should be considered a more superior way to study and understand language? Is the approach indeed *inherently* more rigorous? Or, does its appeal primarily rest on the nature of data?

Finally, here is a question that I would have liked to discuss in person during the candidate's viva hearing (and I do not anticipate that this will be incorporated in the thesis). I am a developmental linguist. My primary interest is the development of language in children. The candidate's questions and his approach to language and data are close to my heart as I am also interested in

'emerging categories'. Some argue that children's language production cannot shed light on the nature of early categories because there is a fundamental difference between children's language use and their underlying language knowledge. What arguments from adult language use and language change would you use to support an emergentist perspective in child language development?

Barbora Skarabela
30 July 2015
Edinburgh, Scotland