

INTERVIEW WITH FLORENT PEREK *ENTREVISTA COM FLORENT PEREK*

*Marcia dos Santos Machado Vieira*¹

*Roberto de Freitas Júnior*²

*Karen Sampaio Braga Alonso*³

ABSTRACT

Professor Florent Perek has a PhD in English and General Linguistics (University of Freiburg, Germany) and is a Lecturer in Cognitive Linguistics at the Department of English Language and Applied Linguistics at University of Birmingham, UK. Professor Perek is the author of several articles in international peer-reviewed journals and has, among his most important publications, the 2015 book, *Argument structure in usage-based construction grammar: experimental and corpus-based perspectives*, edited by John Benjamins.

RESUMO

Florent Perek é Doutor em Inglês e Linguística Geral (Universidade de Freiburg, Alemanha) e Professor da área de Linguística Cognitiva do Departamento de Língua Inglesa e Linguística Aplicada na Universidade de Birmingham, no Reino Unido. Perek é autor de uma série de artigos em artigos em periódicos internacionais revisados por pares e tem, entre suas importantes publicações, seu livro de 2015, o qual foi intitulado *Estrutura argumental na gramática de construções baseada no uso: perspectivas experimental e baseada em corpus* e foi editado pela John Benjamins.

1 Associate Professor at the Vernacular Letters Department (UFRJ) and at the Vernacular Letters Graduate Program (UFRJ). E-mail: marcia@letras.ufrj.br.

2 Adjunct Professor at the Brazilian Sign Language Department (UFRJ), at the Linguistics Graduate Program (UFRJ) and at the Arts and Linguistics Graduate Program (UERJ- FFP). E-mail: robertofrei@letras.ufrj.br.

3 Adjunct Professor at the Linguistics and Philology Department (UFRJ) and at the Linguistics Graduate Program (UFRJ). E-mail: karensampaio@letras.ufrj.br.

We would like to introduce Professor Florent Perek, PhD in English and General Linguistics and Lecturer in Cognitive Linguistics at the Department of English Language and Applied Linguistics at University of Birmingham, UK. Professor Perek is the author of several articles in international peer-reviewed journals and has, among his most important publications, the 2015 book, *Argument structure in usage-based construction grammar: Experimental and corpus-based perspectives*, edited by John Benjamins.

REVISTA LINGUÍSTICA: First, we would like to thank you, Florent, for accepting our invitation for this interview to Revista Linguística, a journal of the Graduate Program on Linguistics of the Federal University of Rio de Janeiro. We would like to begin by asking you to tell us a bit about your most recent works and concerns on language representation, learning and change.

FLORENT PEREK: My earlier research in construction grammar has so far mostly been centered on individual constructions or sets of constructions, with the underlying aim of studying how grammar is acquired, change over time, or simply how it is represented in the mind. This is pretty much what construction grammarians have been doing since the start of the field. However, there is a tendency for a small number of the same constructions to be used over and over in different studies, with comparatively fewer efforts to expand the empirical coverage of construction grammar. This is especially true for argument structure constructions, at least in English, with the ditransitive construction, the caused-motion construction, the resultative construction, the *way*-construction, and a couple of others being systematically used as examples in many studies, including some of my own. There is in principle nothing wrong with that, and furthermore it is perfectly understandable: we do not want to reinvent the wheel and are keen to benefit from an existing body of literature documenting these constructions. But on the other hand, this might give the impression, especially from outside the field, that construction grammar is good at describing a couple of patterns with special properties, but that outside of these, a more mainstream approach like generative grammar would not only be adequate but even preferable, especially for the more mundane and common building blocks of ordinary language (the likes of the transitive construction or the “verb + to-infinitive” pattern, for instance). As construction grammarians, we know that this is not true, but it undermines the theory’s original commitment to account for “the entirety of grammar”. Of course, this commitment was originally meant as “also including the periphery”, as opposed to ‘core’ grammar, but I wonder if construction grammarians have not strayed too far by focusing chiefly on the periphery, as there is also a tendency for us to be ‘butterfly collectors’, i.e. to strive to find new, “exotic” patterns with interest semantic and grammatical properties. However, these are also not common constructions, so it is not clear that focusing on these significantly helps to increase the empirical coverage of construction

grammar, and to demonstrate that it is indeed “constructions all the way down”, as Goldberg puts it.

Motivated by these concerns, in my recent research I have started to embark on a project that aims to expand the empirical coverage of construction grammar by focusing on lesser-studied grammatical patterns, describing them in terms of form-meaning pairs, and trying to catalogue the possible constructions that make up the English language, with a special focus on argument structure constructions, at least initially. In a sense, this is studying constructions in their own right, i.e. as a descriptive exercise that does not in itself aim at a wider goal, at least at first. We believe this is important work that needs to be done, and that it can be of interest not only to other researchers, but also for applications outside academic, for example in language teaching. At Birmingham, my colleague Amanda Pattern and I are currently working towards building a *constructicon* of English, i.e. a database of fully described constructions. In this, we are joining various projects working on constructicons in different languages that are currently in progress around the world; for instance, there is a constructicon of Brazilian Portuguese project led by Tiago Torrent at the Federal University of Juiz de Fora.

REVISTA LINGUÍSTICA: In your webpage, you define yourself as a cognitivist. So, assuming the close relationship between language and cognition, how do you deal with it in your work?

FLORENT PEREK: There are many ways to discuss the relation between language and cognition and how it should apply to language research. To me, it is very aptly captured by Lakoff’s *cognitive commitment*, “a commitment to providing a characterization of general principles for language that accords with what is known about the brain and mind from other disciplines”. In practice, this means that language facts should be analysed and explained with reference to what we know about human cognition, such as categorization, perception, attention, long- and short-term memory, pattern-finding abilities, social cognition, to name but a few. In this light, it is especially useful to look at findings from other fields of cognitive science, in particular psychology (especially cognitive psychology) and neurobiology, but also to some extent artificial intelligence and anthropology. I do, however, see one possible danger in applying the cognitive commitment to the extreme, namely that you run the risk of reading things into your linguistic data by applying some cognitive explanation to it without empirical evidence. To avoid this, it is especially important for linguists to conduct psycholinguistic experiments that directly address cause-effect relations (or at least correlations) between cognition and language.

REVISTA LINGUÍSTICA: Do you agree with the assumption that corpus analysis is a crucial

condition for usage-based linguistic research? If so, do you consider that the analysis of large amounts of data could offer insights about language storage, that is, the psychological reality of grammar?

FLORENT PEREK: It is tempting to create a false equivalence between quantitative corpus linguistics and usage-based linguistics, but in fact not all usage-based linguistics is strictly speaking corpus-based, and not all corpus linguistics necessarily claims to be usage-based in a theoretical sense in the way that it frames its results. Although usage-based linguists have naturally turned to corpus linguistic methods and have done so very early on, we should be conscious of the fact that the use of corpora is not, and has never been, part of the definition of the usage-based approach, as first put forward by Langacker or as later expanded by other users of the term (such as Bybee). The only claim that the usage-based approach makes is that the mental representation of language is shaped by usage events; hence, it is possible to adopt a usage-based approach without quantitative corpus methods. After all, there is ample evidence for the intuitive idea that language users are able to derive linguistic knowledge from a single usage event, with no need for repetition. In this light, you can in principle adopt a usage-based approach by considering individual hand-picked examples, or just on the basis of your own introspection about what possible usage events might be. There is also much experimental research that is squarely usage-based without making reference to corpus data, notably because it only considers usage events within the controlled setting of the experiment; artificial language experiments are a quintessential example of this idea. That said, what corpus data is really good for is telling us how common certain types of usage events are. In my view, that's important for two things: (1) making more reliable generalisations about what aspects of usage are more likely to matter, and (2) investigating frequency effects, i.e. how frequencies of occurrence and co-occurrence shape speakers' knowledge of language.

REVISTA LINGUÍSTICA: How do you see the rise of new forms of digital communication for the creation of new corpora with large amounts of data and for new research questions and methodologies? How can UBCG explore these possibilities?

FLORENT PEREK: I see much value in the data generated by some new forms of computer-mediated communication, such as Twitter, but at the same time I find that they are kind of a double-edged sword: on the one hand, they provide vast amounts of data that are truly unrivaled in size and are public by nature, but on the other hand, the data is extremely specific in its nature, notably in terms of the medium for which it is produced, its intended purpose, and the particular socio-cultural context of use (though admittedly, the same could probably be said of many other specific genres). I find that data from Twitter and other social media is extremely valuable for linguistic research, but

we should be careful what we use it for; there are claims I would be very suspicious of if they were made on the sole basis of a corpus of tweets, however large. That said, I think there is at least one domain in which the value of Twitter data shines through: the study of individual usage, or idiolects, which is a new area of interest in usage-based linguistics. To make any kind of generalizations about the grammar of individual speakers, and compare different individual grammars in a meaningful way, you need a massive amount of data, with clear information as to who produced each text. Twitter provides just that, while traditional corpora are inadequate. The idea is that you should be able to track some aspect of individual speakers' usage from the tweets posted from different accounts. The lines of authorship are sometimes blurred on Twitter (accounts may be shared between different contributors or maintained by a social media expert instead of the account holder), but this is not an insurmountable problem if the data is carefully selected.

REVISTA LINGÜÍSTICA: What are the future challenges for Usage-Based Construction Grammar? What kind of linguistic research (in fields like (supra)segmental phonology, morphology, syntax, semantics and pragmatics) still requires more investigation? What is missing for building bridges with other areas of knowledge and technology?

FLORENT PEREK: I think there are at least two areas where the field is yet to push new boundaries. The first one is the study of individual grammars, as I've already mentioned above. Usage-based approaches to grammar share a strong cognitive basis: they aim to describe grammar as it is 'stored' in the minds of speakers. But on the other hand, they have typically approached grammar in an idealized way, i.e. as a unified and homogeneous entity representative of a whole population of speakers. General-purpose corpora are seen as a way to access the linguistic knowledge of this population. This is in part a necessary fiction created for convenience: we don't aim to describe the knowledge of a single speaker, but rather to make broader generalisations, mostly guided by the assumption that individual speakers must share roughly the same grammar, otherwise communication would be impeded. But if grammar is a cognitive phenomenon, we can't gain a full understanding of it until we take individual minds into account. There are already some studies along these lines, for example focussing on the style of individual authors or speakers and describing/contrasting their use of particular constructions. But we have yet to fully understand how individual usage relates to variation and change in the system as a whole, and for this we need more of this work.

The second one is multimodality in constructions. UBCG is a functional theory; as such it does not study language *per se*, i.e. as a formal system, but as a tool whose primary function is communication. But communication relies on much more than just words, for example gestures

and facial expressions. It is highly desirable to take these other dimensions of communication into account, but it is not clear how they should be integrated with linguistic constructions. To the extent that at least some multimodal signals are meaning-making, are they to be considered constructions in their own right? Or are they additional properties of existing linguistic constructions? Or does their interpretation simply rely on pragmatics rather than conventionalised meanings? To me, the jury is still out, although there have been very valuable contributions to this debate over the past few years. But we still need more research of this kind, especially using quantitative methods, which unfortunately is difficult to do when you're dealing with multimodality.