Calidoscópio
Vol. 15, n. 3, p. 577-580, set/dez 2017
Unisinos - doi: 10.4013/cld.2017.153.16

Frame Semantics and the Texas German dialect research: An interview with Hans C. Boas

Semântica de *Frames* e pesquisa sobre o dialeto alemão-texano: uma entrevista com Hans C. Boas

Profª Drª Rove Chishman
rove@unisinos.br
Universidade do Vale do Rio dos Sinos

Ms. Diego Spader de Souza
dspadersouza@gmail.com
Universidade do Vale do Rio dos Sinos

Prof. Ph.D. Hans C. Boas
hcb@mail.utexas.edu
University of Texas

Interviewers

Dr. Rove Chishman is a Professor in the Applied Linguistics Graduate Program at Universidade do Vale do Rio dos Sinos, RS, Brazil. She holds a PhD in Applied Linguistics by Pontifícia Universidade Católica do Rio Grande do Sul with a post-doctorate at the University of Texas at Austin under Professor Hans Boas' supervision. Her research interests include the interface between Cognitive Semantics theories (with emphasis on Frame Semantics) and Lexicography, Corpus Linguistics and Computational Semantics. She is a full member of the Brazilian National Council of Research (CNPq). Through the coordination of the Research Group SemanTec – Semantics and Technology - Professor Rove Chishman has managed to organize and publish online frame-based dictionaries. Professor Rove currently coordinates two projects: the *Paralympic Dictionary*, which expands on the base of her previous work, and *CNJ-Acadêmico*, which aims at developing an information-retrieval resource for legal language.

Diego Spader de Souza holds a master's degree in Applied Linguistics by Universidade do Vale do Rio dos Sinos, RS, Brazil, and is currently a PhD candidate (with funding by CAPES) at the same institution. He is also a member of SemanTec – Semantics and Technology Research Group - and has worked along Professor Rove Chishman and research colleagues in the development of both dictionaries launched by the group. His research interests include the interface between Frame Semantics and Lexicography, Lexical Semantics, Corpus Linguistics and Computational Linguistics. The PhD dissertation, currently being developed, is concerned with finding how Frame Semantics framework can fit into onomasiology and how it would impact onomasiological lexicography.

Interviewed

PhD Hans C. Boas is an Professor in the Department of German Studies and the Department of Linguistics at the University of Texas at Austin. He holds a PhD in Linguistics by the University of North Carolina at Chapel Hill. His dissertation concerned resultative construction in English and German. In 2011, Prof. Hans was awarded the Leonard Bloomfield Book Award because of the publication of the book called *The life and death of Texas German* in 2009, which resulted of his work with Texas German, an endangered dialect. As an author or editor, he has published several books, chapters and articles in refereed journals of Linguistics. His research interests include Syntax, Lexical Semantics, Computational Lexicography, Language Contact and Variation, Historical Linguistics, Pragmatics, Morphology, Phonology, Documentary Linguistics, Contrastive Linguistics, Corpus Linguistics, Endangered Languages and Dialects, Foreign Language Education, Language Policy and Planning, Intercultural Communication, and History and Philosophy of Linguistics.

Rove Chishman (RC) and Diego Spader de Souza (DSS): Your research has focused on investigating several areas, ranging from studies on the interface between syntax, semantics and pragmatics to linguistic variation and languages in contact, computer lexicons, multilingualism, language policies and legal language. It is also characteristic of your work the dialogue with other fields of study. Taking these aspects into account, how do you see the insertion of your studies in the field of Applied Linguistics?

Hans C. Boas (HCB): I find it difficult to systematically differentiate between "Applied Linguistics" and other fields of linguistics. The study of language has,

[1] Universidade do Vale do Rio dos Sinos. Av. Unisinos, 950, Cristo Rei, 93022-750, São Leopoldo, RS, Brasil.
[2] University of Texas. 110 Inner Campus Drive, Austin, TX 78712, EUA.

in my view, always an "applied" angle (and theoretical insights should in principle be used/applicable, too), because we seek to communicate using language. As such, when studying any type of linguistic phenomenon, I am typically interested in determining how a phenomenon occurs naturally, and how we can use our linguistic insights to solve real-world problems (such as foreign language teaching, intercultural communication, natural language processing of all types, etc.). Of course, that does not mean that we should not think about how to arrive at a theory about how language works, how to represent such knowledge, and ask questions about how this knowledge interacts with other types of knowledge and cognitive processes. But it is important to empirically investigate how language is in fact used, without theoretical straightjackets of various types. I subscribe to usage-based empirical analyses and methodologies in linguistics, i.e. the view that any category, principle, generalization, hypothesis, etc. should in principle be falsifiable by empirical data (scientific principle). Regarding my own studies, I would like to think that, in the various areas such as the syntax-semantics interface, language contact, computational lexicography, etc., they have yielded some insights into our understanding of, for example, how to create more effective electronic resources for (i) learning words in a foreign language, (ii) figuring out how bilingual speakers use (and mix) their languages and why, and (iii) creating an inventory of semantic frames capable of capturing the semantics of lexical units and constructions across different languages.

RC and DSS: In the State of Texas, as well as in the State of Rio Grande do Sul, in Southern Brazil, due to the arrival of German immigrants in the 19th century, there is a strong sociolinguistic context connecting the dialects of German language with the dominant language, be it Portuguese, in the case of Rio Grande do Sul, or English, in the case of Texas. The Texas German Dialect Project, research under your coordination, has as its purposes the preservation and registration of this peculiar linguistic reality. Could you please explain the project and the results already achieved?

HCB: Over the past 16 years, my students and I have recorded more than 600 speakers of Texas German, a critically endangered immigrant dialect that has been spoken in central Texas since the 1840s. In about 20 years this unique dialect will be extinct. The recordings consist of open-ended sociolinguistic interviews and translations of lists of words, phrases, and sentences. The recordings, together with their metadata, are stored in a freely available online archive (www.tgdp.org). So far we have transcribed several hundred hours of the more than one thousand hours of recordings. The archived materials are used for teaching, outreach, and research. We use the recordings, together with their transcriptions, in our linguistics classes at UT Austin to engage our students with naturally occurring data. We also use the recordings to create teaching materials for classes on culture, history, and immigration. One thing that I never thought would happen was that a composer took the transcripts of our recordings to create the "Texas Liebeslieder", describing the experience of the German immigrants and their descendants, in the style of Johannes Brahms' 19th century "Liebeslieder" (love songs). Finally, we use the archived materials to conduct research on the various aspects of language contact, presenting our results at academic conferences, and publishing them as articles and books.

RC and DSS: One of the interfaces established by Cognitive Linguistics is that with Computing, especially considering the development of the Berkeley FrameNet Project (https://framenet.icsi.berkeley.edu/) and the consequent development of similar lexical databases for other languages, as it is the case of German (http://www.laits.utexas.edu/gframenet/), Spanish (http://spanishfn.org/), Japanese (http://jfn.st.hc.keio.ac.jp/) and Brazilian Portuguese (http://www.ufjf.br/framenetbr/). In your opinion, what are the positive aspects of adopting Frame Semantics for the construction of these multilingual computational resources?

HCB: Frame Semantics is one of the few theories about word meaning that has actually been applied to the creation of a large corpus-based lexical database, FrameNet, at the International Computer Science Institute in Berkeley. Over the past 20 years, the FrameNet team has created an impressive amount of lexical entries for English verbs, nouns, adjectives, adverbs, and prepositions, using semantic frames to structure the lexicon of English. But since much of the workflow of FrameNet is manual, it takes a lot of time to cover large amounts of data, which means that the coverage of FrameNet is not as great as that of other lexical databases (e.g. WordNet). However, the amount of detailed information in FrameNet entries is exhaustive, allowing users (and programs) to learn more about the meanings of words (and other words sharing that meaning) and how those meanings are realized syntactically.

FrameNet has been used by a variety of NLP applications. Perhaps one of the biggest and in my view most exciting results is that the semantic frames created on the basis of English can be re-used for creating similar resources for other languages. There are some typological and culturally-specific differences that need to be accounted for, but so far it looks as if about 90%, possibly more, of the semantic frames based on English can be re-used for accounting for the lexicons of other languages (so far only a small set of about a dozen, so we might end up with different results after covering more languages). As such, semantic frames seem to allow for a systematic comparison of the lexicons of different languages, and they can also serve as meaning anchors for developing so-called constructicons (databases containing construction entries) for different languages, too.

RC and DSS: One of the most debated issues among cognitive linguists today concerns research methodology. In your opinion, what are the central aspects of this debate? And what is the status that Corpus Linguistics has received in Cognitive Linguistics studies?

HCB: This is a very tricky question, because if we really want to subscribe to empirical research in linguistics, then Corpus Linguistics is only part of the answer. Cognitive Linguistics, like other theories, often makes strong claims about how language works at a cognitive level. Many studies provide fascinating insights and proposals, and they might be right, but at the end of the day, we often don't know what is really going on in our brains and bodies, i.e. how we learn, store, and process language (of course, this holds for other theories of language, too). For that we would need to know a lot more about what is going on in our brains, i.e. we need to have empirical data that can be measured and used for building hypotheses that could be tested.

Right now we are at the beginning of a very exciting time because the technical machinery such as brain imaging allows us for the first time to systematically study what is going on in the brain when we "do things" with

language. But until we have a more complete picture of how the brain/the mind (and the body) really do things with language, I am hesitant to make definite claims about the cognitive reality of theoretical insights in linguistics (including Cognitive Linguistics). In that respect, the current situation in linguistics can be compared with the study of black holes in outer space. Astronomers have studied them quite extensively relying on different sources of data and as a result of their observations they have been able to state a series of hypotheses and theories about the nature of black holes. These can be tested against existing data and new data down the road. But no one has actually traveled into a black hole to collect data on what is going on inside a black hole (or beyond). Our current state of knowledge leads us to claim that traveling into a black hole is impossible because we would be crushed by the gravity in it. As such, the types of statements about what is going on inside a black hole are only hypotheses which so far cannot be tested using data from inside the black hole. In present-day linguistics the situation is roughly similar. We can use all sorts of data to arrive at hypotheses and theories about most aspects of the nature of language (including the cognitive reality of what is going on in our brains when we use language), but until we have the technical machinery to more fully understand how the brain does things with language, it is, in my view, very difficult to really empirically test these claims.

However, the situation is not as gloomy as it seems. The past 25 years have seen the emergence of Corpus Linguistics as an empirical way to study language. Before the 1990s, a lot of research in linguistics relied on intuitions and anecdotal evidence to arrive at hypotheses and theories about how language works. Some people have labeled this approach to the study of language as "armchair linguistics" or "speculative linguistics". But the creation of large-scale linguistic corpora of different types and the tools to explore them has allowed linguists to develop specific empirical methodologies to search for specific patterns and relationships between units of various granularities. These methodologies make use of, among other things, frequencies and statistical correlations in order to help us better identify and measure specific linguistic phenomena so that we can develop hypotheses, which in turn can be tested using the scientific method. The emergence of Corpus Linguistics is very exciting, but, in my view, one should also pay attention to not overemphasize the role of statistical information in Cognitive Linguistics. The last decade or so has seen a number of detailed studies on a variety of linguistic phenomena, but there has, in my view, been too often an overemphasis on statistical patterns and correlations to explain the linguistic phenomenon under investigation. In my view, corpus linguistics and the associated statistical methodologies are very important and necessary to help us understand the nature of linguistic phenomena, but they typically are not strong enough to provide us with linguistic explanations per se.

RC and DSS: Considering your works in the field of Cognitive Linguistics, Construction Grammar and Frame Semantics are among your main theoretical references, could you talk about your interest in such approaches and how they can be integrated?

HCB: Given the intellectual history of the field, I would say that both Frame Semantics and Construction Grammar are direct descendants of Fillmore's seminal (1968) paper "The Case for Case". After abandoning case grammar during the 1970s, Fillmore started developing his theory of Frame Semantics in the late 1970s and early 1980s. In parallel, Fillmore and his colleagues sought to develop an alternative theory of language that cared about the entirety of language, not only certain aspects of it (core vs. periphery/competence vs. performance/separation between syntax and other modules of language, etc.). This theory, which started with a series of in-depth case studies of semi-idiomatic grammatical constructions, came to be known as Construction Grammar during the mid-1980s. During the 1990s and beyond, construction grammarians paid a lot of attention to so-called argument structure constructions, then word order constructions and many other different types of constructions beginning in the early 2000s.

The important point to remember is that Construction Grammar and Frame Semantics are sister theories. They share the notion that language consists of a structured inventory of linguistic signs, pairings of form and meaning. "Form" in this sense is to be interpreted as lexical, syntactic, morphological, phonetic/phonological, etc. while "meaning" is to be interpreted as (lexical) semantics, pragmatics, etc. Looking at the concept of construction, its meaning can, in most cases, be represented using semantic frames. This means that when we study constructions, we are also almost always dealing with semantic frames. I regard the two sister theories of Frame Semantics and Construction Grammar as a part of the larger enterprise of Cognitive Linguistics, where, I think, many if not most linguistic insights are compatible with each other.

When it comes to the applied angle of Cognitive Linguistics, it is important to note that Frame Semantics is the theoretical prerequisite for FrameNet. Without Frame Semantics, there would be no FrameNet. The types of semantic frames found in FrameNet, the lexical entries, and other information could be regarded as "applied" Frame Semantics. But it does not stop there. Over the past few years, several research groups have started developing so-called "constructicons", databases consisting of entries of grammatical constructions that are an extension of (lexical) FrameNet. As such, these constructicons can be seen as "applied" Construction Grammar and as a natural extension of "applied" Frame Semantics as in FrameNet. In short: there has already been an integration of the two theories of Construction Grammar and Frame Semantics at both the theoretical and applied levels.

RC and DSS: It is often said that Cognitive Linguistics is not a theory, but an archipelago of theories that share common commitments. Another aspect that draws attention is the interdisciplinarity with other sciences. Considering such amplitude, could one say that Cognitive Linguistics has the same object of study? What view of cognition supports such a diversity of studies?

HCB: I find it problematic to identify specific aspects of such studies to determine what types of investigation contribute to our larger understanding of cognition and language. At this point, I think we are still in the exploration phase, similar to the 13th century when certain parts of the world were not well known to people living in Europe and they first had to discover what can be found on the planet, then how to explore it and map it. One example is multimodal Construction Grammar, which has been growing very strong over the past few years. Again, this sub-field has had tremendous first successes largely because of the availability of corpora that

allow us to systematically study multimodal phenomena. From my perspective, it is clear that such interactions of different types of linguistic and other types of cognitive information must be investigated in great detail, but it is, at the present point, too early to come to any definite conclusions.

RC and DSS: After forty years of its emergence, Cognitive Linguistics has currently been diffusing very expressively in the world and reaching a significant number of adepts. In addition to the United States and Europe, Cognitive Linguistics today has branches in Asia and South America. Considering this vast scenario, what are the main themes of Cognitive Linguistics today?

HCB: I don't know. I guess it depends on who you ask because most people I know have different interests in exploring different types of themes in Cognitive Linguistics. What I would say is that the common assumptions are probably important, i.e. that language is a part of a larger cognitive apparatus, that language is embodied, that language is not necessarily modular, etc.

RC and DSS: Finally, considering that Applied Linguistics is today open to a variety of different types of research regarding language, culture, society and interaction, how do you see the future of Applied Linguistics and what do you expect from the future generations of researchers in this field?

HCB: Given my own experience with Construction Grammar and Frame Semantics and how they have been successfully applied to the creation of FrameNets and constructicons for different languages, I would personally like to see three main things accomplished over the next five decades or so. First, it would be exciting to greatly expand the coverage of the Berkeley (English) FrameNet (or a FrameNet for some other language). The proof of concept is already there, but expanding FrameNet to cover 150,000 lexical units or more is going to be very expensive. Should that hold us back from doing it? I believe not. To come back to my comparison between space exploration and linguistics: In the early 1960s, the United States made a pledge to send a man to the moon by the end of the decade, and they did it. Yes, it was expensive, but it could be done given the proper financial support. I regard a possible major future expansion of FrameNet similarly. Second, it would be exciting to create a full-coverage constructicon of English (or some other language), creating construction entries for all the constructions and linking them in structured networks so that the information can be used to investigate a wide range of linguistic phenomena more systematically. Finally, it would be great to do this not just for one language, but preferably for a multitude of languages. I am aware that these expectations might be somewhat unrealistic, given the dire funding situation, but this should not stop us from thinking and dreaming about them. If engineers are capable of sending people to the moon, linguists should be able to systematically document, explore, and analyze human language in its entirety.