



# Introduction of Prof. Dr. Ulrich Heid

## Plenary speaker at the EUROPHRAS2015 Conference

Malaga, Spain, 29 June 2015

by [Violeta Seretan](#), University of Geneva

Good evening, ladies and gentlemen, and welcome to the second plenary session of the EUROPHRAS 2015 Conference. My name is Violeta Seretan and it is my very great pleasure and honour to introduce our speaker: One of the leading researchers in our field; a pillar of computational linguistics and a founder of modern, computational phraseology; a world class researcher who had the opportunity and also the responsibility of shaping this field and this research area which brought us all here in Malaga.

Indeed, virtually all of us know his work because we are all interested in the *ways words work together*. His '94 paper with the same title is a reference in many fields, including Lexicography, Terminology, Second Language Learning, Natural Language Processing (NLP), and it is one of the most cited papers in his endless list of publications that he produced in German, English, or French since he started his career in the late '80s.

He studied Romance Philology (French and Italian) and history, but happily for our field in '86 he got interested in computational linguistics, lexicography and terminography, having developed a special affinity for lexical affinities, or collocation, which is actually the topic of his talk. (N. B. *Extracting linguistic knowledge about collocations from corpora.*)

He received his PhD in Computational Linguistics from Universität Stuttgart in '95 for a thesis on *structuring monolingual and contrastive electronic dictionaries*. He was a post-doc at ICSI, Berkeley in the FrameNet group, then he came back to Stuttgart, where he got his "habilitation" in 2001 and worked at the Institute for NLP, ever since the Institute was founded and until a few years ago.

He is most known for his work on phraseology, but his activity was in many areas of language processing – or in German *Sprachverarbeitung*; Institute for NLP is in German *Institut für Maschinelle Sprachverarbeitung, IMS*. At IMS he led (or took part in) many projects, national and international on topics as varied as: machine translation and the reuse of lexical and terminological resources (in Eurotra-7); infrastructure and standardisation (in ISLE, EAGLES, and D-SPIN which is the German CLARIN), dialogue systems, multi-modality, annotation; or again, computational morphology, information extraction, lexical semantics, disambiguation, in so many projects it is hard to count.



## Computerised and Corpus-based Approaches to Phraseology: Monolingual and Multilingual Perspectives

29 June to 1 July 2015



Closer to our interests, one of his most well-known achievements is the corpus query tool *IMS Corpus Workbench* which includes *CQP/CQL*, the *corpus query processor/language*. This is the output of a project from '94 and today it is still a standard, a reference support in corpus based lexicography, terminology and in corpus linguistics and over the years it has served as a basis for many developments including *CQPweb* and *BNCweb*. This is just one example, but the list of achievements he made and research topics he worked on are so long I simply cannot enumerate. To put it briefly, our field owes a lot to him.

In 2010, he has been appointed full professor of language technology and computational linguistics at Universität Hildesheim. In this capacity, he is working on computational linguistics support to digital humanities, specialised terminology, and the design of a new age of electronic dictionaries which should be more than simple digital copies of paper dictionaries. He has a vast experience in collaborating with the industry and using language technologies for corpus based dictionary building.

Another recent major achievement is the co-editing of the fourth volume of the HSK handbooks: *Dictionaries. An International Encyclopedia of Lexicography*. This is an inventory of the advances in lexicography in the past decades, focused on electronic and computer-assisted lexicography. This 16 hundred page volume completes the 3 volumes edited in '89-'91 by Franz Josef Hausmann and colleagues and is a monumental work destined to be the reference for dictionary research. He also acts as an editor of the journal *Lexicographica* and the accompanying book series.

He influenced our field not only through his own work, but also through the work of the researchers he supervised. To mention just a few: Fabienne Cap, Marion Weller, Denis Spohr and his thesis on computational lexicon modelling, or again Stefan Evert and his thesis on lexical association measures, well-known in NLP. At the time Stefan Evert was finishing his thesis, I was beginning mine in collocation extraction in Geneva, and I visited IMS in Stuttgart. In the office of our guest I remember there was a whiteboard on which he had written a couple of words together with some features: number, determination, and so on. It was in 2004, and it was the beginning of a new series of experiments, a new idea that he was popularising in the years to come, and pursuing in projects, as well as using in collaborations with the industry: the idea that with today's language technologies we should be able not only to extract collocations from corpora, but also to extract detailed linguistic information about them for lexicographic purposes.

But to what extent this can be achieved? And why does it matter in text production, from a monolingual and from a multilingual perspective? This is precisely what our guest will tell us. So ladies and gentlemen, without further ado, please welcome our distinguished speaker, Prof. Dr. Ulrich Heid!