

COMPUTATIONAL LINGUISTICS - AN INTERVIEW WITH RENATA VIEIRA

Renata Vieira

Universidade do Vale do Rio dos Sinos – UNISINOS

ReVEL – What is the field of Computational Linguistics concerned with? What is the focus of language that Computational Linguistics aims at studying?

Renata - The field is concerned with the understanding of the language by itself and of the appropriate computational techniques for the treatment of written and spoken language for both its interpretation and generation.

ReVEL – The field of Natural Language Processing (NLP) is relatively new. How it began to get structured? Which were its pioneering researchers and projects?

Renata – It is possible to say that this field has nearly half a century. It began along with the field of Artificial Intelligence (AI) which aims at reproducing intelligent behavior in computer systems, such as problem solving and automated reasoning. A very popular example of AI's application is the development autonomously playing games, like Chess. Since the beginning of AI, one of the features of intelligence investigated by researchers was the human ability of communication. Research development on this

issue led to the generation of the field called NLP. The first projects of PLN were related to automatic translation.

ReVEL - Among the many problems concerning PLN's research, which are considered more complex to you? What is the state-of-the-art in this field?

Renata – There are so many problems that would be difficult to enumerate them. Most of the problems are related to ambiguity, references to the context and to the difficulty of representing the common sense knowledge - all of these aspects play an important role on the interpretation of the language. However there were great advances in this field, such as the recognition and the speech synthesis which are technologies almost mastered by researchers but still with some restrictions. I will take this case to exemplify some of the complex problems related to this area. Such a system is capable of recognizing the speech of a person from his/her interaction with it, but without proper training the same system can not recognize the speech of a second speaker. For an independently performed recognition of the speaker the vocabulary to be recognized must be very restricted.

ReVEL - In which field of the Computational Linguistics we can perceive a greater development in recent years?

Renata – In addition to the example mentioned above, a more recent trend in this area is to move from a small experimental set towards working with large-scale problems, such as extensive linguistic corpora. The field of syntax analysis is well developed and established. The field of computational semantics has also evolved a lot, but once it is a more challenging field there are still many unsolved problems.

ReVEL – What books would you suggest for those who are just starting their studies in Computational Linguistics?

Renata – Most books in this field are written in English. The book of James Allen, *Natural Language Understanding*, Addison Wesley Publishing Company, 1995, can be considered as canonical literature. There are other more recent books - very interesting as textbooks - but with a more computational approach, like: *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics and Speech Recognition*, Dan Jurafsky, James H. Martin, Keith Vander Linden, Nigel Ward, Daniel Jurafsky Prentice Hall, 2000, and *Foundations of Statistical Natural Language Processing*, Christopher D. Manning, MIT Press, 1999. There is also a forthcoming book edited by myself and by Professor Vera Lucia Strube de Lima, from PUCRS, with contributions from many Brazilian scholars of the field of NLP. This book will be published this year by Editora Manole.