NITheCS

**National Institute for
Theoretical and Computational Sciences**

NITheCS COLLOQUIUM:

Describing language: formal and empirical approaches

Prof Menno van Zaanen (South African Centre for Digital Language Resources)

Monday, 27 February 2023 | 16h00 – 17h00 SAST

Venue: online

ABSTRACT

Natural languages (those we speak as humans) have certain regularities and restrictions. For instance, we cannot simply put words together and expect to end up with valid sentences. If we want to describe which sequences lead to valid language, we need to decide on a formalism that allows for such descriptions. The main aim is then to identify a language formalism that can identify exactly which sequences are considered valid and which invalid in a natural language.

In this presentation, I will provide some examples of language formalisms. To start, I introduce a few simple formalisms with known mathematical properties. If we try to describe natural languages with these formalisms, it quickly becomes clear that more complex formalisms are required to properly describe natural languages. More complex formalisms, however, also require more computational complexity (i.e., it will take more effort to decide whether a sequence is valid or not). At the same time, we can show that even simple formalisms are not learnable in an efficient manner. This clashes with the idea that humans are quite capable of learning natural languages relatively efficiently. I will briefly propose different ways of tackling this apparent paradox (by revisiting the language formalisms and the notion of efficient learnability).

Finally, I will relate this to practical descriptions of natural languages.

BIOGRAPHY

As a professor in Digital Humanities, Menno is particularly interested in incorporating the use of computational techniques in the field of Humanities. His PhD in the area of computer science dealt with building systems that learn (linguistic) grammars from plain sequences (sentences). These empirical grammatical inference systems result in patterns that can be used for further analysis of the data, for instance, in applied machine learning, computational linguistics, or computational musicology. During his MA (computational linguistics) and MSc (computer science) studies, Menno used techniques from the one field and applied it to situations in the other, such as proofing tools and error correction, machine translation, and multi-modal information retrieval. Such techniques can be applied to Humanities data, but for them to be fully successful, the results still need to be interpreted in the context of Humanities.

REGISTER
TO ATTEND

Visit https://bit.ly/3Xm231Z
or scan/click:

**MEET THE SPEAKER
AFTER THE EVENT**
visit:
kumospace.com/nithecs_social

For more events, visit www.nithecs.ac.za