## Asoc.prof. Sanda Martinčić-Ipšić, University of Rijeka, Croatia

## Title: Language Networks

**Abstract:** Written, as well as spoken language can be modelled via complex networks where the lingual units (words) are represented by nodes and their linguistic interactions by links. Such representations enable language analysis through varying linguistic units. So far there have been efforts to model isolated phenomena of various language subsystems and examine their unique function through complex networks, although failing to explain the mechanism of their mutual interactions. Obtaining such findings may be critical for deepening our understanding of conceptual universalities in natural languages, especially in shedding light on the cognitive representation of a language.

Recently, the focus of complex networks' research has shifted from the analysis of isolated properties of a system toward a more realistic modeling of multiple phenomena - multilayer networks. The multilayer network of language is a unified framework for modeling linguistic subsystems and their structural properties enabling the exploration of their mutual interactions. Various aspects of natural language systems can be represented as complex networks, whose vertices depict linguistic units, while links model their relations. The multilayer network of language is defined by three aspects: the network construction principle, the linguistic subsystem and the language of interest. More precisely, we construct a word-level (syntax and co-occurrence) and a subword-level (syllables and graphemes) network layers, from four variations of original text (in the modeled language). The analysis and comparison of layers at the word and subword-levels is employed in order to determine the mechanism of the structural influences between linguistic units and subsystems. The obtained results suggest that there are substantial differences between the networks' structures of different language subsystems, which are hidden during the exploration of an isolated layer. The word-level layers share structural properties regardless of the language (e.g. Croatian or English), while the syllabic subword-level expresses more language dependent structural properties. The preserved weighted overlap quantifies the similarity of word-level layers in weighted and directed networks.

The complex networks framework can be utilized in applications as well. Our approach proposes a novel network measure - the node selectivity for the task of keyword extraction. The node selectivity is defined as the average strength of the node. Firstly, we show that selectivity-based keyword extraction slightly outperforms the extraction based on the standard centrality measures: in-degree, out-degree, betweenness, and closeness. Furthermore, from the data set of Croatian news we extract keyword candidates and expand extracted nodes to word-tuples ranked with the highest in/out selectivity values. The obtained sets are evaluated on manually annotated keywords: for the set of extracted key-word candidates. Selectivity-based extraction does not require linguistic knowledge as it is derived purely from statistical and structural information of the network. The experimental results point out that selectivity-based keyword extraction has a great potential for the collection-oriented keyword extraction task.

CV: Sanda Martinčić-Ipšić obtained her B.Sc. degree in Computer Science in 1994 from the University of Ljubljana Faculty of Computer Science and Informatics, and her M.Sc. degree in informatics from the University of Ljubljana, Faculty of Economy in 1999. In 2007 she obtained a Ph.D degree in Computer Science from the University of Zagreb, Faculty of Electrical Engineering and Computing. Dr. Martinčić-Ipšić currently works as an associate professor of Computer Science at the University of Rijeka, Department of Informatics. Her research interests include speech and language technologies, natural language processing, complex networks, automatic speech recognition, speech synthesis, corpora development, with a special focus on the Croatian language. She has published more than 60 scientific papers and books.