From an enormous Excel sheet with 24000 Croatian verbs to the newest idea about creating a comprehensive resource with max. 1000 Croatian verbs with semantic roles attached

I intend to present a personal insight of an expert associate, i.e. my ten-year experience with resources dealing with verb valences at the Institute of Croatian Language and Linguistics. Although I was initially employed as an expert associate at the corpora *Croatian Language Repository*, that component of the work on computational linguistics at the Institute of Croatian Language and Linguistics was stopped by the sudden departure of the head of the department (dr. sc. Damir Ćavar), and the department for computational linguistics itself was abolished. After several years of working on national terminological projects within the program *Development of Croatian Special Field Terminology – Struna*, at the end of 2013, I joined (as the only member with any computational linguistics experience) another team that was doing preliminary work for the creation of the Valency Dictionary of Croatian Verbs. Their previous work was focused exclusively on the printed dictionary experiences, so this team didn't even have a clearly defined data structuring model or a diagram for the dictionary verb lemmata entry and processing.

Therefore, I want to show how, without much programming/IT support or a computational linguistics department, we created a schematization of data types and relations for the Valency Base of Croatian Verbs and then also for the publication of the first part of the e-Glava valence dictionary online, after months of studying previously unstructured and scattered documents and questioning colleagues who theoretically dealt with verb valences. Dictionary entries were organized according to the German valency tradition. They include semantic description, not through semantic roles, but rather through argument definitions. In the meantime, another Institute team was completing the project *Croatian Web Dictionary - Mrežnik* (free, monolingual, easily searchable hypertext online dictionary of the Croatian standard language with 10,000 entries). Within those 10,000 lemmas, about 1,000 frequent verbs in contemporary Croatian were processed. Although those entries don't include classic semantic roles either, there is a part of their description with the arguments and circumstances of predicate structures, extracted according to the most common collocations in the corpus.

In the final part of the presentation, I would like to present a fresh idea about (semi)automatic conversion of these various semantic descriptions (from two traditional dictionaries) into universal tags with semantic roles, thus creating a new, modern valence dictionary as a digital resource with at least 500 to 700 verbs in Croatian with semantic roles attached. I would also like the final part of the analysis

to serve as a stimulus for sharing your ideas and experiences with similar resources, i.e. how you think I could organize the process the best.