Hungarian determiners from a computational linguistic point of view

Noémi Ligeti-Nagy
MTA-PPKE Hungarian Language Technology Research Group
ligeti-nagy.noemi@itk.ppke.hu

Tag
Datum
Zeit
Raum

In this paper, we attempt to describe an algorithm to handle determiners in Hungarian sentences within the framework of a psycholinguistically motivated parsing system (*AnaGramma*, Prószéky-Indig 2005).
We collected a list of determiners and the morphological annotation of these words from our corpus (Pázmány Corpus, Endrédy 2016). We distinguished two types of determiners: the ones already annotated as determiners (definite articles and three archaic words) and the ones tagged as pronouns (or numerals). In the current state of *AnaGramma*, the feature DET is a supply, and every case ending releases a demand, looking for a DET. If there is not any, the demand is fulfilled with default mechanisms. This method works well with articles as they bear the feature DET in their annotation. However, it is not working with other determiners, as they are simply tagged as PRON (pronoun), NUM (numeral) etc. This is the problem needed to be solved: how can we manage this kind of determiners to make them able to be a supply demanded by case endings of noun phrases; and in the same time, keep their ability to be subjects, objects etc. of a verb (which is not a feature of the members of the other group of determiners). To resolve this contradiction, we suggest that these tokens should, later on, be annotated based on their original part-of-speech; nevertheless, to make them able to act as determiners, namely to provide a supply being able to fulfill a demand, an extra DET feature should be added to their annotation.
After implementing the algorithm and extending the annotation of the determiners, we shall be able to handle a great amount of noun phrases correctly, which may lead to a better performance of *AnaGramma*.

References: • Prószékéky, G. and Indig, B. (2015): Magyar szövegek pszicholingvisztikai indíttatású elemzése számítógéppel [Psycholinguistically motivated parsing of Hungarian texts]. *Alkalmazott nyelvtudomány* 15, 39–44 (Original document in Hungarian). • Endrédy, I. (2016): Nyelvtechnológiai algoritmusok korpuszok automatikus építéséhez és pontosabb feldolgozásukhoz [Language technology algorithms for automatic corpus building and more precise data processing]. PhD dissertation. PPKE-ITK (Original document in Hungarian).