

## Automatic Error Detection concerning the Definite and Indefinite Conjugation in Texts by Learners of Hungarian

Veronika Vincze<sup>1</sup>, János Zsibrita<sup>2</sup>, Péter Durst<sup>3</sup>, Martina Katalin Szabó<sup>4</sup>

<sup>1</sup>Hungarian Academy of Sciences, Research Group on Artificial Intelligence  
vinczev@inf.u-szeged.hu

<sup>2</sup>University of Szeged, Department of Informatics  
zsibrita@inf.u-szeged.hu

<sup>3</sup>University of Szeged, Hungarian Studies Center  
durst.peter@gmail.com

<sup>4</sup>University of Szeged, Hungarian Linguistics PhD Programme  
szabomartinakatalin@gmail.com

In this paper we focus on automatic error detection concerning the definite and indefinite conjugation in Hungarian, based on data from the HunLearner corpus.

The texts of HunLearner were POS-tagged and dependency parsed by *magyarlanc*, a linguistic preprocessing toolkit of Hungarian. On the basis of the syntactic and morphological analysis we were able to define rules for the object-verb agreement, which made it possible to collect those sentences where there was a mismatch between the definiteness of the object and the verbal conjugational pattern.

Here we just focused on cases where the object is phonologically present in the sentence, so we neglected cases when the presence of the pronominal object could be only deduced from the verbal form. We also neglected cases when the object was a subordinate clause.

Our results reveal grammatical structures that might pose problems for learners of Hungarian. The most frequent source of errors was when the object is a common noun with a definite article: it triggers definite conjugation but in 17% of the errors, it co-occurred with an indefinite verb. Other frequent errors are a demonstrative pronoun as the object and a bare common noun (i.e. without an article) as the object: in 13-13% of the errors, they do not co-occur with the required type of conjugation. Together with the errors induced by possessive forms, these types altogether are responsible for 50% of the mismatches in conjugation.

It is also shown that the definite object + indefinite conjugation (59%) is a much more frequent phenomenon than the opposite, i.e. indefinite object + definite conjugation.

Our results may be fruitfully applied in language teaching on the one hand as the statistical analysis makes it possible for the students to concentrate on grammatical structures that seem to give rise to more difficulties. On the other hand, from a natural language processing point of view, definiteness errors in conjugation may be automatically corrected as the automatic detection of the type of the object triggers the type of conjugation. If the sentence does not contain the required form, a grammar checker may automatically propose some corrections concerning the word form of the verb.