On the Importance of Ezafe Construction in Persian Parsing

Alireza Nourian, Mohammad Sadegh Rasooli, Mohsen Imany and Heshaam Faili

Data Preparation

- We attach the Ezafe indicator to the POS tags and train a sequence tagger on the new tagset.

- We have also manually Ezafe tagged all words in the Persian dependency treebank with 99.6% annotator agreement.

- We define some rules to convert a dependency tree to a shallow phrase structure. Implementation of the these rules is available in the Hazm toolkit:

https://github.com/sobhe/hazm

Python library for digesting Persian text.

Acknowledgements

We thank Computer Research Center of Islamic Sciences (CRCIS) for supporting us on corpus annotation.