

Overview of the SPMRL 2013 Shared Task: Cross- Framework Evaluation of Parsing Morphologically Rich Languages

Central topic

- Provide standard datasets for morphologically rich languages in different representations and parsing scenarios.
- Standardize the evaluation protocol on morphologically ambiguous input
- Raise community awareness with regard to the difficulty of parsing morphologically rich languages

Methodology

Datasets

- Include data in both constituency and dependency annotation.
- *full data* setup and *small* setup (5,000 sentences)
- Three parsing scenarios:
 - gold segmentation, pos tags, and morphological features are provided
 - automatically predicted segmentation, pos tags and features

- lattice of multiple possible morphological analyses and joint disambiguation of the morphological analysis and syntactic structure

Findings

Previous research

- first statistical parsing models were generative and based upon treebank grammars

“ applying the phrase-based treebank grammar techniques is sensitive to language and annotation properties, and these models are not easily portable across languages and schemes.

Notable quotes

“ While progress on parsing English -- the main language of focus for the ACL community -- has inspired some advances on other languages, it has not, by itself, yielded high-quality parsing for other languages and domains. This holds in particular for morphologically rich languages... where important information concerning the predicate-argument structure of sentences is expressed through word formation, rather than constituent-order patterns as is the case in English and other configurational languages. p. 146

“ recently, advances in PCFG-LA parsing (Petrov et al. 2006) and language-agnostic data-driven dependency parsing (McDonald et al. 2005; Nivre et al. 2007b) have made it possible to reach high accuracy with classical feature engineering techniques in addition to, or instead of, language specific knowledge. p. 147

Follow up readings

Revision #2

Created Tue, Jan 14, 2020 12:59 AM by [kenneth](#)

Updated Tue, Jan 14, 2020 1:36 AM by [kenneth](#)