Yoav Goldberg

Computer Science, Ben Gurion University of the Negev

Dealing with the Complexities of Syntactic Parsing in Hebrew:
Addressing agreement, word-segmentation and rich morphology in a fast dependency parser

I will describe my experience with designing a syntactic parser for Hebrew, a language with rich morphology and a small treebank. After describing some of the characteristics that make automatic syntactic processing of Hebrew challenging and discuss some data representation issues, I will present some solutions to these challenges. These include improvements of a semi-supervised broad-coverage tagger, and a greedy dependency parser which can accommodate rich feature-sets and cope with noisy data while remaining fast. I will also briefly discuss a constituency parsing system that performs joint morphological segmentation and syntactic parsing. The work on Hebrew brought about solutions that work well also for English; I will point to these results when appropriate.

Yoav Goldberg is finishing up his PhD in Ben-Gurion university in Israel under the supervision of Prof. Michael Elhadad. His PhD work revolves around natural language processing in Hebrew and morphologically rich languages, syntactic parsing, and doing fun stuff with computers and language. Prior to his PhD he participated in the Israeli Hi-Tech industry, mostly as a freelance security consultant. His research interests include structured prediction, computational creativity and confidence estimation.

http://www.cs.cmu.edu/~nlp-lunch/