

Lexical Adaptation of Link Grammar to the Biomedical Sublanguage: a Comparative Evaluation of Three Approaches

Sampo Pyysalo and **Tapio Salakoski** **Sophie Aubin** and **Adeline Nazarenko**
Turku Centre for Computer Science (TUCS) LIPN
and University of Turku Université Paris 13 & CNRS UMR 7030
Lemminkäisenkatu 14 A, 99, av. J.-B. Clément,
FIN-20520 Turku, Finland F-93430 Villetaneuse, France

Abstract

We study the adaptation of Link Grammar Parser to the biomedical sublanguage with a focus on domain terms not found in a general parser lexicon. Using two biomedical corpora, we implement and evaluate three approaches to addressing unknown words: automatic lexicon expansion, the use of morphological clues, and disambiguation using a part-of-speech tagger. We evaluate each approach separately for its effect on parsing performance and consider combinations of these approaches. In addition to a 45% increase in parsing efficiency, we find that the best approach, incorporating information from a domain part-of-speech tagger, offers a statistically significant 10% relative decrease in error. The adapted parser is available under an open-source license at <http://www.it.utu.fi/biolg>.
