

of rule sorting and rule movements. For most parameter constellations, repetition of optimization runs did not lead to a better performance than a single pass, unless each iteration is allowed to add new sections. The first-pass average improvement in 5-fold cross evaluation was 0.31 percentage points (F-Score 96.17), similar to Danish results reported in Bick (2013), but with added sectionizing and long iterations, improvements of over 1 percentage point were seen, corresponding to a 30% improvement in relative terms. The immediate effect was best for precision, but with high iterations, recall was affected most, with a 60% improvement in relative terms.

Future work would certainly profit from access to a large computer cluster, to investigate the millions of possible combinations of incremental parameter changes. Also, it would be interesting to get the human linguist back into the loop, to see if some of the rules slated for killing or demotion by the optimizer can be "saved" by additional context conditions instead, and if the best selected generalized variants of wordform rules can be used for further development.

References

- Bick, Eckhard. 2013. ML-Tuned Constraint Grammars. In: Proceedings of the 27th Pacific Asia Conference on Language, Information and Computation, pp. 440-449. Taipei: Department of English, National Chengchi University.
- Bick, Eckhard. 2014. ML-Optimization of Ported Constraint Grammars. In: Calzolari, Nicoletta et al. (eds.), Proceedings of the 9th International Conference on Language Resources and Evaluation, LREC2014 (Reykjavik, May 28-30, 2014), pp. 3382-3386.
- Bick, Eckhard & Didriksen, Tino. 2015. CG-3 - Beyond Classical Constraint Grammar. In: Proceedings of NoDaLiDa 2015 (forthcoming).
- Faarlund, Jan Terje, Lie, Svein & Vannebo, Kjell Ivar. 1995. Norsk referansegrammatikk. Oslo: Universitetsforlaget.
- Hagen, Kristin & Nøklestad, Anders. 2010. Bruk av et norsk leksikon til tagging og andre språkteknologiske formål. *LexicoNordica* 2010 (17) pp. 55-72.
- Hagen, Kristin & Johannessen, Janne Bondi. 2003. Parsing Nordic Languages (PaNoLa) - norsk versjon. Nordisk Sprogteknologi 2002. Museum Tusulanums Forlag, Københavns universitet.
- Johannessen, Janne Bondi and Helge Hauglin. 1998. An Automatic Analysis of Norwegian Compounds. In Haukioja, T. (ed.): Papers from the 16th Scandinavian Conference of Linguistics, Turku/Åbo, Finland 1996 : 209-220.
- Johannessen, Janne Bondi, Hagen, Kristin & Nøklestad, Anders. 2000. A Constraint-based Tagger for Norwegian. In 17th Scandinavian Conference of Linguistics [Odense Working Papers in Language and Communication 19].
- Johannessen, Janne Bond; Hagen, Kristin; Lynum, André; Nøklestad, Anders. 2012. OBT+stat: A combined rule-based and statistical tagger. In Andersen, Gisle (ed.). Exploring Newspaper Language: Using the web to create and investigate a large corpus of modern Norwegian, s. 51–66.
- Karlsson, Fred, Voutilainen, Atro, Heikkilä, Juha & Anttila, Arto. 1995. Constraint Grammar: A Language-Independent System for Parsing Unrestricted Text. In Natural Language Processing, No 4. Berlin and New York: Mouton de Gruyter.
- Lager, Torbjörn. 1999. The μ -TBL System: Logic Programming Tools for Transformation-Based Learning. In: Proceedings of CoNLL'99, Bergen.
- Lindberg, Nikolaj & Eineborg, Martin. 1998. Learning Constraint Grammar-style Disambiguation Rules using Inductive Logic Programming. COLING-ACL 1998: 775-779
- Norsk Ordbank 'Norwegian Word Bank'. 2010. <http://www.hf.uio.no/iln/om/organisasjon/edd/forskning/norsk-ordbank/>.
- Oslo-Bergen Tagger homepage. <<http://tekstlab.uio.no/obt-ny/>>.