

Norwegian Dependency Treebank with c-structure

The Language Bank

This is a version of the Norwegian dependency treebank (NDT) 2.0 where constituent structure (c-structure) has been added. The original version of NDT can be downloaded from <https://www.nb.no/sprakbanken/ressurskatalog/oai-nb-no-sbr-86/>

In order to add constituency, the same idea as in Vinyals et al. (2015) has been operationalized by fine-tuning the large north T5 model (https://huggingface.co/north/t5_large_NCC) with the labelled bracketed sentences for nno-ndt-lfg and nob-ndt-lfg, found at <https://clarino.uib.no/iness-prod/sentences>. The sentences in both nno-ndt-lfg and nob-ndt-lfg have been parsed and disambiguated in the Norwegian LFG treebank using the NorGram LFG grammar.

Acknowledgements

We acknowledge INESS (<https://clarino.uib.no/iness>) for the use of NorGramBank to create this treebank.

Victoria Rosén, Koenraad De Smedt, Paul Meurer, and Helge Dyvik. [An open infrastructure for advanced treebanking](#). In Jan Hajič, Koenraad De Smedt, Marko Tadić, and António Branco (eds.) *META-RESEARCH Workshop on Advanced Treebanking at LREC2012*, pages 22–29, Istanbul, Turkey, May 2012.

If you have questions or comments, please feel free to contact us at sprakbanken@nb.no.

References

Vinyals, Oriol, Lukasz Kaiser, Terry Koo, Slav Petrov, Ilya Sutskever, and Geoffrey E. Hinton. 2015. Grammar as a foreign language. In *Advances in Neural Information Processing Systems 28: Annual Conference on Neural Information Processing Systems 2015, December 7-12, 2015, Montreal, Quebec, Canada*, pp. 2773–2781. URL <http://papers.nips.cc/paper/5635-grammars-a-foreign-language>.

Norsk dependenstrebant (NDT) med c-struktur

Språkbanken

Dette er en versjon av Norsk dependenstrebant (NDT), versjon 2.0, hvor det er lagt til konstituentstruktur (c-struktur). Den originale versjonen av NDT kan lastes ned fra <https://www.nb.no/sprakbanken/ressurskatalog/oai-nb-no-sbr-86/>

For å legge til c-struktur har ideen i Vinyals et al. (2015) blitt operasjonalisert ved å finjustere den store North-T5-modellen (https://huggingface.co/north/t5_large_NCC) med setninger i nno-ndt-lfg og nob-ndt-lfg fra INESS (se <https://clarino.uib.no/iness-prod/sentences>). Setningene i både nno-ndt-lfg og nob-ndt-lfg har blitt parset og disambiguert i den norske LFG-trebanken ved bruk av LFG-grammatikken NorGram.

Takk

Takk til INESS (<https://clarino.uib.no/iness>) for bruken av NorGramBank for å lage denne trebanken.

Victoria Rosén, Koenraad De Smedt, Paul Meurer og Helge Dyvik. 2012. [An open infrastructure for advanced treebanking](#). I Jan Hajič, Koenraad De Smedt, Marko Tadić og António Branco (red.) *META-RESEARCH Workshop on Advanced Treebanking at LREC2012*, side 22–29, Istanbul, Turkey, May 2012.

Referanser

Vinyals, Oriol, Lukasz Kaiser, Terry Koo, Slav Petrov, Ilya Sutskever, and Geoffrey E. Hinton. 2015. Grammar as a foreign language. I *Advances in Neural Information Processing Systems 28: Annual Conference on Neural Information Processing Systems 2015, December 7-12, 2015, Montreal, Quebec, Canada*, side 2773– 2781.
URL: <http://papers.nips.cc/paper/5635-grammaras-a-foreign-language>