

Site report for Trondheim 2021

Maintained applications, grammar developments, valence, and Ga

Lars Hellan

DELPH-IN Summit, July 2021

Contents

Updates by June 30, 2021:

- Maintained applications
- *Norsource* developments
- Norwegian valence resources
- Small grammar for Ga

Maintained applications - technical development and maintenance by *Tore Bruland*

Norwegian Grammar Sparrer, a 'mal'-grammar built on Norsource (2013 -):

<http://regdili.hf.ntnu.no:8081/studentAce/parse>

Grammar demo for Norsource (2013 -):

<http://regdili.hf.ntnu.no:8081/linguisticAce/parse>

Norwegian Valence Corpus, produced from Norsource parsing (2017 -):

https://typecraft.org/tc2wiki/Norwegian_Valency_Corpus

Multilingual valence lexicon (*MultiVal*), hosting Bulgarian, Ga, Norwegian and Spanish, based on synchronized verb classifications from the respective grammars (2014 -):

http://regdili.hf.ntnu.no:8081/multilanguage_valence_demo/multivalence

Tagger based on Norsource (new version):

<http://regdili.hf.ntnu.no:8081/bonTaggerApp/tag>

References concerning standing applications

- Hellan, L., Bruland, T., Aamot, E., Sandøy, M.H. (2013): A Grammar Sparrer for Norwegian. *Proceedings of NoDaLiDa2013*.
- Hellan, L., D. Beermann, T. Bruland, M.E.K. Dakubu, and M. Marimon(2014) *MultiVal: Towards a multilingual valence lexicon*. In Calzolari et al. (eds) 2014.
- Hellan, Lars and Tore Bruland. 2015. A cluster of applications around a Deep Grammar. In: Vetulaniet al. (eds) *Proceedings from The Language & Technology Conference(LTC) 2015, Poznan*.
- Hellan, Lars, Dorothee Beermann, Tore Bruland, Tormod Haugland, Elias Aamot. 2020. Creating a Norwegian valence corpus from a deep grammar. In *Human Language Technology. Challenges for Computer Science and Linguistics. 8th Language & Technology Conferene, LTC2017*. Springer 2020 ISBN 978-3-030-66526-5; web demo: https://typecraft.org/tc2wiki/Norwegian_Valency_Corpus).

Extending Norsource

The main effort related to *Norsource* over the last year has been extending the number of verb entries by nearly 3000, the number of verb entries now being around 15,600.

This has been aligned with a completion of the verb valence catalogue *NorVal*, described in later slides.

A few rules have been added to the grammar, accounting for the free variation in order between pronominal object and particle, significantly increasing coverage and simplifying lexicon.

Technically, most of the applications mentioned above are hosted at a web server at NTNU (*'regdili'*). The server environment is being updated, and Tore has installed the applications on a new machine, taking effect this summer.

Developing Norwegian valence resources - NorVal

This application goes back to the *TROLL* project in the late 80ies. The TROLL lexicon, extended as the *NorKompLeks* lexicon, played a major role in the first development of Norsource in 2001-02.

Hellan, Lars, Lars G. Johnsen and Anneliese Pitz. 1989. TROLL. University of Trondheim. (Downloadable at Nasjonalbiblioteket.)

Nordgård, Torbjørn. 1998. "Norwegian Computational Lexicon (NorKompLeks)". Proceedings of NoDaLiDa 1998.

The continued development of Norsource in turn provides a basis for the creation of a free-standing valence resource, called **NorVal**, first presented in the works below:

- Hellan, Lars. 2021a. A valence catalogue for Norwegian. To appear in: *Natural Language Processing in Artificial Intelligence - NLPinAI 2021*. Springer Series Studies in Computational Intelligence (SCI); <https://www.springer.com/series/7092>
- Hellan, Lars. 2021b. Norwegian Valence Dictionary (NorVal). Paper to be presented at SLE (Societas Linguistica Europaeae) 2021

Developing Norwegian valence resources - NorVal

The formal connection between an HPSG grammar formalism and the classification system used in NorVal is outlined in the following paper:

- Hellan, Lars. 2019. Construction-Based Compositional Grammar. *Journal of Logic Language and Information*. DOI: 10.1007/s10849-019-09284-5.

Prospects for how to extend a system like that of NorVal to include senses are discussed in the following (and in Hellan 2021c on next slide):

- Hellan, Lars. 2020a. Interoperable Semantic Annotation. LREC workshop ISA-16, 6th Joint ACL-ISO Workshop on Interoperable Semantic Annotation

Developing Norwegian valence resources - NorVal

Among linguistic papers dealing with crucial aspects of valence frame identification in Norwegian are:

- Hellan, Lars. 2020b. Representing Light Reflexives in a valence resource for Norwegian. Presentation at SLE 2020.
- Hellan, Lars. 2021c. Unification and selection in Light Verb Constructions. A study of Norwegian. To appear in *Light verb constructions as complex verbs. Features, typology and function* (Series “Trends in Linguistics. Studies and Monographs”), Mouton de Gruyter.)
- Hellan, Lars and Dorothee Beermann. 2020. Presentational and related constructions in Norwegian with reference to German. In: Abraham, Werner, Elisabeth Leiss and Yasuhiro Fujinawa (eds.) *Thetics and Categoricals*. [LA 262] , John Benjamins Publishing Company.

The Ga computational grammar

The work with an HPSG grammar for the West African language Ga, started in 2005 but for a long time discontinued, was resumed and described in the following paper:

Lars Hellan. 2020d. A computational grammar of Ga. 1st Workshop on Resources for African Indigenous Languages (RAIL) , LREC 2020. | [workshop web site](#) | [Proceedings \(pdf\)](#) | [online workshop on May 16, 2020](#)

For download of files:

https://typecraft.org/tc2wiki/Ga_computational_HPSG_grammar

The work with this grammar has also provided a source for the development of lexical resources for Ga, most notably valence resources, described in the above paper. With the combined efforts with Mary Esther Dakubu and Tore Bruland we created:

https://typecraft.org/tc2wiki/Ga_Valence_Profile

The Ga computational grammar

These resources , together with the following,

- Dakubu, Mary Esther Kropp. Unpublished a. 'Ga_verb_dictionary_for_digital_processing'. Accessed for download at: https://typecraft.org/tc2wiki/Ga_Valence_Profile
- Dakubu, Mary Esther Kropp. Unpublished b. Ga Verbs and their constructions. Monograph ms, Univ. of Ghana.

remain to be brought to a formal and publication stage like that of NorVal.

The Ga computational grammar

The value of such resources in the development of methods for NLP for so-called low resourced languages should not be underestimated. Cf.

- *Beermann, Dorothee; Hellan, Lars.*
Enhancing grammar and valence resources for Akan and Ga. In: *West African languages. Linguistic theory and communication*. Warszawa: Wydawnictwa Uniwersytetu Warszawskiego 2020 ISBN 978-83-235-4623-8. s. 166-185
- *Beermann, Dorothee; Hellan, Lars; Mihaylov, Pavel; Struck, Anna.*
Developing a Twi (Asante) Dictionary from Akan Interlinear Glossed Texts. In: *Proceedings of the LREC 2020 1st Joint SLTU and CCURL Workshop (SLTU-CCURL 2020)*. European Language Resources Association 2020 ISBN 979-10-95546-35-1. s. 294-297

Hellan, Lars.

How Aspects of Descriptive and Formal Linguistics Can Inform LT for All Languages.
UNESCO conference ,Paris: *Language Technology For ALL*; 2019-12-04 - 2019-12-06

End

Thanks, these were the updates till end of June 2021.