Site report Trondheim

NorSource related work

Lars Hellan and Tore Bruland

DELPH-IN Summit, July 2019, Cambridge

Over the last year

- Increasing coverage relative to *Personalhåndboka*, accommodating legal-style long sentences and construction types

- Developing more robust strategies of parsing (Tore; will tell about it next year)

- Further work on a TFS-based frame for language typology (TypeGram; later today)

- Developing comparative valence resources (for Wednesday)

- Maintaining existing applications (mainly through monitoring, since they run well by themselves)

- Preparing for grammar documentation, given an attained balance between increased coverage and retained analytic depth

Applications running (with steadily increased use)

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 automatically 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

TypeGram, a package for grammar induction based on TypeCraft IGT (2011-):

https://typecraft.org/tc2wiki/TypeGram

Grammar Induction

Our package for Grammar Induction from IGT with valency annotation (see Hellan (2010), Bruland (2011), and Hellan and Beermann (2011, 2014))

is downloadable from:

http://regdili.hf.ntnu.no:8081/typegramusers/menu

and also via – where an introduction is given:

http://typecraft.org/tc2wiki/TypeGram

NorSource Webdemo Statistikk

Version 1.0

År	Demo	Antall
2019	tag	29
2019	stud	82225
2019	mlv	269
2019	ling	739
2018	tag	100
2018	stud	121157
2018	mlv	450
2018	ling	1860
2017	tag	209
2017	stud	92272
2017	mlv	485
2017	ling	1375
2016	tag	349
2016	stud	67983
2016	mlv	448
2016	ling	1636
2015	tag	231
2015	stud	59631
2015	mlv	588

År	Måned	Demo	Antall
2019	7	tag	1
2019	7	stud	4165
2019	7	mlv	17
2019	7	ling	16
2019	6	tag	7
2019	6	stud	11781
2019	6	mlv	42
2019	6	ling	71
2019	5	tag	1
2019	5	stud	11910
2019	5	mlv	30
2019	5	ling	74
2019	4	tag	9
2019	4	stud	12748
2019	4	mlv	51
2019	4	ling	88
2019	3	tag	1
2019	3	stud	14164
2019	3	mlv	44

Example of batch of 10 sentences with responses in English:

Norwegian Grammar Tutor

Demo with ACE, version 1.11. For further guidelines, see Info

Enter up to 10 sentences with up to 10 words each and press the Analyze button.

Feedback Language [English ▾]

```
Du liker jeg.
Jeg prøver og komme.
Ola å Per kommer.
Jeg skammer seg.
Imorgen jeg kommer.
Kommer jeg snart.
Jeg like fisken.
Jeg prøvde å gikk.
Husen er gult.
En gult bil stod her.
```

[Analyze]

[generate]	The word "jeg" is marked with the wrong case, try using "meg" instead. More description
[generate]	The word "og" is not the infinitival marker, try using "å" instead. More description
3	**Ungrammatical in Norwegian**
[generate]	The reflexive pronoun "seg" does not match the number and gender of the word it refers back to. Try using "meg" instead. More description
[generate]	The sentence lacks subject-verb inversion. More description
[generate]	The sentence contains an incorrect subject-verb inversion. More description
[generate]	The word "like" has the wrong tense.. More description
8	**Ungrammatical in Norwegian**
[generate]	The word "husen" is of neuter gender, not masculine. More description
[generate]	The adjective "gult" is conjugated as neuter gender, but modifies a masculine or feminine noun. More description

Example of batch of 10 sentences with responses in Polish:

Norwegian Grammar Tutor

Demo with ACE, version 1.11. For further guidelines, see Info

Enter up to 10 sentences with up to 10 words each and press the Analyze button.

Feedback Language [polski ▼]

```
Du liker jeg.
Jeg prøver og komme.
Ola å Per kommer.
Jeg skammer seg.
Imorgen jeg kommer.
Kommer jeg snart.
Jeg like fisken.
Jeg prøvde å gikk.
Husen er gult.
En gult bil stod her.
```

[Analyze]

generate	Wyraz "jeg" jest w nieprawidłowym przypadku, spróbuj użyć "meg". More description
generate	Wyraz "og" nie jest markerem bezokolicznikowym, spróbuj użyć "å". More description
3	**Ungrammatical in Norwegian**
generate	Zaimek zwrotny "seg" nie zgadza się z liczbą i rodzajem wyrazu, do którego się odnosi. Spróbuj użyć "meg". More description
generate	W tym zdaniu brakuje inwersji podmiot-czasownik. More description
generate	To zdanie zawiera nieprawidłową inwersję (odwróconą kolejność) podmiot-czasownik. More description
generate	Wyraz "like" jest w nieprawidłowym czasie. More description
8	**Ungrammatical in Norwegian**
generate	Wyraz "husen" jest rodzaju nijakiego, a nie męskiego. More description
generate	Przymiotnik "gult" odmienia się jak w rodzaju nijakim, ale określa rzeczownik męski lub żeński. More description

Using the button '*Generate*' to see acceptable option

Norwegian Grammar Tutor

Demo with ACE, version 1.11. For further guidelines, see Info

Enter up to 10 sentences with up to 10 words each and press the Analyze button.

Feedback Language polski ▼

```
Du liker jeg.
Jeg prøver og komme.
Ola å Per kommer.
Jeg skammer seg.
Imorgen jeg kommer.
Kommer jeg snart.
Jeg like fisken.
Jeg prøvde å gikk.
Husen er gult.
En gult bil stod her.
```

 Analyze

Grammar Option(s) for Sentence

#	Sentence
1	Jeg prøver å komme .

Grammar Sparrer web pages, *with number of views until now*

- Main intro page [Classroom:Norwegian Grammar Checking](#)   251,437
- [Feedback messages](#) 8,954
- [Korektor gramatyczny dla języka norweskiego](#) ([Ćwiczenia z gramatyki norweskiej z polskim komentarzem](#))   11,758
- [Informacje zwrotne](#)   1,344
- [挪威语语法检测软件](#)   4,483
- [Controllo della grammatica norvegese](#) 2,826
- [Messaggi di feedback](#)   3,056
- [Rückmeldungen](#)   2,447

- Pages describing phenoena of Norwegian grammar, views from 20,000 to 2,000:

Texts accompanying the Grammar Sparrer

Related pages

[The Noun Phrase - Norwegian](#)

[Agreement in Norwegian noun phrases](#)

[Definite determiners in Norwegian](#)

[Possessive constructions in Norwegian](#)

[Gender in Norwegian nouns](#)

[Coordination marking in Norwegian](#)

[Sentence syntax - Norwegian](#)

[Subject-Verb Inversion in Norwegian](#)

[Sentence adverbials in Norwegian](#)

[Verb Complementation - Norwegian](#)

[Infinitives in Norwegian](#)

[Past and Perfective patterns in Norwegian](#)

[Personal pronouns in Norwegian](#)

[Reflexives - Norwegian](#)

[Reflexive verbs in Norwegian](#)

[Verb - Preposition expressions in Norwegian](#)

Webdemo analysis format, also for text files.
Ex: "jeg skrubber Ola på ryggen" 'I scrub Ola on his back'

head-subject-rule (0,5) -0.193830
  jeg_perspron (0,1) 0.000000
    jeg
  head-actv-verb-icomp-rule (1,5) -0.169794
    head-verb-comp-rule (1,3) 0.260878
      pres-infl_rule (1,2) -0.635503
        skrubbe_detachposs_vlxm (1,2) -0.219792
          skrubber
      Ola-firstname_pn (2,3) 0.000000
        ola
    head-prep-comp-rule (3,5) -0.133957
      på_detachposs (3,4) -0.183730
        på
      sg_def_m_final-full_irule (4,5) 0.384731
        sg-masc-def-noun-lxm-lrule (4,5) 0.000000
          rygg_relinan_nlxm (4,5) 0.000000
            ryggen

Ex: "jeg skrubber Ola på ryggen". MRS

(saying essentially that Ola has a back, both of which are involved in the scrubbing)

ltop=h0, index=e1

h3:_1sg-pron_rel([ARG0:x2])

h4:_pronoun_q_rel([ARG0:x2, RSTR:h5, BODY:h6])

h7:_skrubbe_v-trObl_rel([ARG0:e1, ARG1:x2, ARG2:x8, ARGOBLQ:h9])

h10:_possess_rel([ARG0:e11, ARG1:x8, ARG2:x12])

h13:named_rel([CARG:Ola, ARG0:x8])

h15:_def_q_rel([ARG0:x8, RSTR:h16, BODY:h17])

h9:_på_p-possrais_rel([ARG0:u18, ARG1:e1, ARG2:x12, IARG:u19])

h20:_rygg_n_rel([ARG0:x12])

h21:_def_q_rel([ARG0:x12, RSTR:h22, BODY:h23])

< qeq(h5,h3), qeq(h16,h13), qeq(h22,h20) >

Typecraft Valence + IGT   'normal form'

String: Jeg vet at hun forbauset Ola
Free translation: I know that she surprised Ola

Morph	Jeg		vet		at		hun		forbause		t		Ola
Citation			|vite					|forbause				
		1.SG.NOM		PRES		DECL		3.SG.FEM				PRET		
	|PN	|V	|COMP	|PN	|V				|Np

vet: SAS: NP+Sdecl
          FCT: transWithSentCompl
          ConstructionLabel: v-tr-obDECL

forbauset: SAS: NP+NP
                 FCT: transitive
                 ConstructionLabel: v-tr

From 'deep' grammar to corpus

Valence corpora have often been built manually, or by statistical methods where hand annotation plays a crucial role. In some cases valence corpora, possibly in conjunction with tree-banks, are used in the construction of computational grammars. Here we go the opposite way, exporting information from a deep grammar to an IGT corpus, whereby sentences in the corpus serve as classified examples of the verb valence types of the language.

- English: FrameNet, VerbNet and PropBank (http://verbs.colorado.edu/~mpalmer/projects/verbnet.html),

- German: Evalbu (http://hypermedia2.ids-mannheim.de/evalbu/);

- Czech Vallex (http://ucnk.ff.cuni.cz) ;

- Polish, Walenty (http://clip.ipipan.waw.pl/Walenty; Przepiórkowski & al (2014),

- Cf., Osenova, (2011), Patujek and Przepiórkowski (2016).

References

- Beermann, Dorothee and Mihaylov, Pavel, 2014. Collaborative databasing and Resource sharing for Linguists. *Languages Resources and Evaluation* 48. Dordrecht: Springer, 1-23.

- Beermann, Dorothee, Lars Hellan and Tormod Haugland: Convergent development of digital resources for West African Languages. *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)* (CCURL eds.: Claudia Soria, Laurent Besacier and Laurette Pretorius)

- Dakubu, M.E. Kropp and Lars Hellan, 2017. A labeling system for valency: linguistic coverage and applications. In Hellan, L., Malchukov, A., and Cennamo, M (eds) *Contrastive studies in Valency*. Amsterdam & Philadelphia: John Benjamins Publishing Co.

- Hellan, Lars and M.E. Kropp Dakubu, 2010. *Identifying verb constructions cross-linguistically*. In *Studies in the Languages of the Volta Basin* 6.3. Legon: Linguistics Department, University of Ghana.

- 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, L., Bruland, T., Aamot, E., Sandøy, M.H. (2013): A Grammar Sparrer for Norwegian. *Proceedings of NoDaLiDa* 2013.

- Hellan, Lars and Tore Bruland. 2015. A cluster of applications around a Deep Grammar. In: Vetulani et al. (eds) *Proceedings from The Language & Technology Conference* (LTC) 2015, Poznan.

- Hellan, Lars, Dorothee Beermann, Tore Bruland, Tormod Haugland, Elias Aamot. 2017. Creating a Norwegian valence corpus from a deep grammar. In *Proceedings from LTC 2017*, Poznań, edited by Zygmunt Vetulani. 425-429. Poznan : Adam Mickiewicz University