

Site report - *Summer 2016:*

NorSource related work

in the *Research Group in Digital Linguistics, NTNU, Trondheim*
(where ***TypeCraft*** is the other main application, instrumental also
in the applications to be shown)

Lars Hellan, Tore Bruland, Elias Aamot

presented by Dan Flickinger

DELPH-IN Summit, June 16, 2016

Stanford

NorSource – *phases as of last year*

The grammar since its start in 2002 has gone through the following phases:

- Phase 1, the *Grounding* phase (2001-04),
- Phase 2, the *Semantic Expansion* phase (2005-07),
- Phase 3, the *Cross-Linguistic Coding* phase (2008-10), and
- Phase 4, the *Applications* phase (2010-).
- Web site: http://typecraft.org/tc2wiki/Norwegian_HPSG_grammar_NorSource
- As for Phase 2: a tdl-file with the semantics of spatial and temporal relations for prepositions and adverbs, and sample lexical entries, fitting into any current Matrix grammar, can be downloaded from the site.

As for Phase 4: Two main applications ... :

Application 1. Constructing an e-learning tool from an LKB grammar

The ***Norwegian Online Grammar Sparrer*** is an online language training tool developed at NTNU, with a direct access point at

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

and a wiki access point at

[http://typecraft.org/tc2wiki/A Norwegian Grammar Sparrer](http://typecraft.org/tc2wiki/A_Norwegian_Grammar_Sparrer)

An introduction to its ‘mal-grammar’-based design is given in Hellan et al. 2013. This application, and a derived ‘Scrabble’ game, will be the topic of this presentation.

Application 2. **MultiVal** – the Multilingual Valency database

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

Guidelines: http://typecraft.org/tc2wiki/Multilingual_Verb_Valence_Lexicon

The system hosts 4 languages, with altogether 50 000 verb entries, with valency frames classified in a uniform system. The languages hosted are:

Bulgarian (*lexicon import from BURGER, the Bulgarian Matrix grammar*)

Ga (*lexicon import from GaGram, the Ga Matrix grammar, whose lexicon is in turn imported from a ToolBox lexicon of Ga, created by M.E.Kropp Dakubu*)

Norwegian (*lexicon import from NorSource , the Norwegian Matrix grammar*)

Spanish (*lexicon import from SRC , the Spanish Matrix grammar*)

For documentation per February 2014 (before Bulgarian got added), see Hellan et al., LREC 2014.

The web system is stable, and steadily used, but not massively. What such a resource needs in addition is systematic semantic information. (Addressing this aspect, we, i.a., work, TypeCraft and MultiVal jointly, on the connection with *ImagAct*.)

NorSource related

A REST API design has been created whereby more applications can be run on NorSource.

First of these is a ***Sentence Scrabble***, which builds on the Grammar Sparrer, but is designed as a 'put sentences together from words' game, with 100 words available for each game.

Illustrations follow, but first an update on the Sparrer itself.

The Sparrer

As before the Sparrer accepts *batches of up to 10 sentences*, each with max. 10 words. The design with ***freely chosen inputs*** is maintained, which, with a lexicon of nearly 85,000 entries and a grammar with 250 syntax rules etc., can never be under full control; but it's improving...

Responses are now given (in order of creation) in ***English, Polish, Italian, German, Chinese, Bulgarian, Arabic*** , and ***Norwegian*** (all languages supplied collegially and for free, from friends in in- and outland).

Pages with examples of feedback messages are also available on TypeCraft for most languages, while the wiki pages describing aspects of Norwegian grammar – now about 20 pages – are still only in English.

Explanations can help only so much, so considerable work has been spent during the last year on making the ***Generation*** facility more robust, so that by now, nearly all of the 130 mal-messages are accompanied by generation of a/the correct version (see later).

Next to the Sparrer – Norwegian Sentence Scrabble

<http://regdili.hf.ntnu.no:5051/page/#/play>

Combines the grammar sparrer concept with ***gamification*** to create a more engaging user experience.

Uses the same parsing, generation and error message components as the grammar sparrer through the aforementioned API.

100 words, from a total set of 1100, are provided for each game, in ratios between nouns, verbs, adjectives and the various more functional word classes, alphabetically ordered in one list. The system uses the full set of syntax rules, but a restricted vocabulary, which lets us avoid the most problematic word/construction combinations.

Instruction on the game site:

Use the provided words to form grammatical Norwegian sentences. Nouns, adjectives and verbs are provided in their basic form. These have to be inflected when put together to sentences. Other words can only be used in the form provided.

Refresh the browser to start a new game.

Next to the Sparrer – Norwegian Sentence Scrabble

The Scrabble is not yet ‘advertised’, as we are still in a developing phase, and with little time.

A minor improvement to be made is to order nouns by themselves, verbs and adjectives by themselves, to make it easier to look for a ‘plot’ for each sentence.

A major desideratum will be to somehow distinguish between semantically reasonable and semantically less reasonable sentences (short of that, one can use words as ‘ontologically neutral’ as possible).

And related to this, explore vocabulary size and composition for optimal engagement and learning.

A further problem, also for the Sparrer, is how to reward grammatically complex sentences without leading the users beyond the capabilities of the grammar.

With grammar mistakes ...

The system is based on the same set of error handling and feedback messages as the Grammar Sparrer, illustrated below. In such cases the number of error messages is subtracted from the score otherwise earned by the number of words;

2 points: den tørr akademiker sove

The adjective "tørr" is conjugated as an indefinite, but stands with a definite determiner.

The word "sove" is in infinitive, but should be in past or present tense.

Suggested correction: Den tørre akademikerer sover

Next slides: a game at initial stage, the view at an ended game, and a game started with feedback in Chinese.

Norwegian Grammar Scrabble - points 7

aktivitet	altfor	annerledes	at	av	be	behagelig	billig	bli	bli
bo	borte	de	den	den	der	dere	din	ei	eller
en	eng	far	farlig	fire	fjell	fordi	fortelle	fotball	fram
gate	gjette	gjette	gutt	ha	ham	heller	hjelp	hjemme	holde
hotell	hun	hva	hvilken	hvorvidt	håpe	håpe	høres	intet	kald
kalle	kaste	kg	lege	lese	ligge	med	meg	min	måned
noen	nokså	nær	eg	og	også	om	palestinsk	pike	plen
pølii	regne	se	seg	sende	sjelden	skamme	sko	skulle	slik
smyge	sekk	som	sove	spise	spise	stole	storm	stå	syk
så	tillate	tre	utenfor	vaske	vinter	vær	være	være	øre

Enter sentence here

CHECK

END GAME

4 points: Sokken tillater ham og lese.

- The word "og" is not the infinitival marker, try using "å" instead.

Suggested correction: *Sokken tillater ham å lese .*

3 points: Politiet smyger seg.

Use the provided words to form grammatical Norwegian sentences. Nouns, adjectives and verbs are provided in their basic form. These have to be inflected when put together to sentences. Other words can only be used in the form provided.

Refresh the browser to start a new game.

For comments and feedback, contact elias.aamot@gmail.com

High Score

Lars H	60
	44

- START NEW GAME
- START NY RUNDE (NORSK)
- START NEW GAME (ARABIC)
- START NEW GAME (POLISH)
- START NEW GAME (ITALIANO)
- NEUES SPIEL (DEUTSCH)
- START NEW GAME (ZHONGWEN)
- START NEW GAME (BULGARIAN)



Norwegian Grammar Scrabble - points 5

alle	ansettelse	anta	antagelig	barn	be	bero	bestemor	bli	burde
bygd	de	de	den	dere	det	du	dusin	dusin	ei
en	fire	fjord	flat	foreløpig	fortelle	få	gram	ha	ha
han	hans	heller	hennes	holde	hvem	hvilken	hvis	hyggelig	håpe
i	kinesisk	koke	kort	lav	legge	legge	lese	love	lure
løpe	mann	meg	meget	mens	min	mulig	måtte	ned	noen
nokså	og	og	om	pike	politi	prøve	reserve	selv	sende
skamme	slem	sne	sol	soldat	sparke	sprekke	spørre	spørs	storm
syng	syrisk	så	så	søndag	ta	tilbake	time	to	under
unne	utenfor	varm	vaske	vei	vokter	vår	vær	være	å

Enter sentence here

Alle kinesisk pikene måtte leser den.

CHECK

END GAME

3 points: Alle kinesisk pikene måtte leser den.

- 形容词 "kinesisk" 是一个不定指变体，但是这里却和一个定指限定词连用。
- 如果一个定指名词被一个形容词修饰，例如 "kinesisk"，那么这个形容词前面应该有一个限定词。
- 单词 "leser" 是现在式，应该用不定式。

Suggested correction: *Alle de kinesiske pikene måtte lese den .***2 points: Han være slem.**

- 单词 "være" 是不定式，应该用过去式或现在式。

Suggested correction: *Han er slem .*

Use the provided words to form grammatical Norwegian sentences. Nouns, adjectives and verbs are provided in their basic form. These have to be inflected when put together to sentences. Other words can only be used in the form provided.

Refresh the browser to start a new game.

For comments and feedback, contact elias.aamot@gmail.com

Glimpses from the Sparrer

Returning to the Sparrer, we first show a batch of four sentences, all ungrammatical in ways pointed out in the ma-lmessages, and then examples of what gets generated in three of the cases.

After that, for the multi-lingual pleasure of it, the mal-responses as they look for some other languages, and finally the feedback overview, in one of the languages. (Although there are 130 token mal-messages, they are grouped in 25 main types, and one example per type is offered.)

The Sparrer – illustrating a small batch, responses in English

Norwegian Grammar Tutor

Demo with ACE, version 2.2 for further guidelines, see [Info](#)

**Enter up to 10 sentences with up to 10 words each and press the Analyze button.
Always use a verb and a subject in your sentences**

Feedback Language English ▾

```
mannen smile  
et gutt ikke kommer  
jeg ser gutten sin  
Ola skammer
```

Analyze

generate

The word "smile" is in infinitive, but should be in past or present tense. [More description](#)

generate

A determiner must have the same gender, number and definiteness as the noun it modifies. [More description](#)
An adverb cannot precede the finite verb in a main clause.

generate

The reflexive pronoun "sin" does not match the person of the word it refers back to. Try using "min" instead. [More description](#)

generate

The verb "skammer" requires a reflexive object. [More description](#)

Generation for sentence 2 – “*et gutt ikke kommer”

Norwegian Grammar Tutor

Demo with ACE, version 2.2 for further guidelines, see [Info](#)

**Enter up to 10 sentences with up to 10 words each and press the Analyze button.
Always use a verb and a subject in your sentences**

Feedback Language

```
mannen smile  
et gutt ikke kommer  
jeg ser gutten sin  
Ola skammer
```

Analyze

Grammar Option(s) for Sentence

#	Sentence
1	En gutt kommer ikke

Generation for sentence 3 – “*Jeg ser gutten sin”

Norwegian Grammar Tutor

Demo with ACE, version 2.2 for further guidelines, see [Info](#)

**Enter up to 10 sentences with up to 10 words each and press the Analyze button.
Always use a verb and a subject in your sentences**

Feedback Language

```
mannen smile  
et gutt ikke kommer  
jeg ser gutten sin  
Ola skammer
```

Analyze

Grammar Option(s) for Sentence

#	Sentence
1	Jeg ser gutten min



Norwegian Grammar Tutor

Demo with ACE, version 2.2 for further guidelines, see [Info](#)

**Enter up to 10 sentences with up to 10 words each and press the Analyze button.
Always use a verb and a subject in your sentences**

Feedback Language Deutsch ▼

```
mannen smile  
et godt ikke kommer  
jeg ser gutten sin  
Ola skammer
```

Analyze

generate

Das Wort "smile" ist im Infinitiv, muss aber eine Vergangenheits- oder Gegenwartsform annehmen. [More description](#)

generate

Der Artikel muss die selben Genus-, Nummer- und Definitivspezifikationen haben wie das Nomen. [More description](#)
Das Adverb kann in Hauptsätzen nicht dem finiten Verb vorausgehen.

generate

Das possessive Reflexivpronomen "sin" steht nicht in Kongruenz (Person) mit dem Wort, auf das es sich bezieht.
Benutzen Sie "min". [More description](#)

generate

Das Verb "skammer" ist ein reflexives Verb und muss daher immer zusammen mit einem passenden Reflexivpronomen benutzt werden. [More description](#)

Norwegian Grammar Tutor

Demo with ACE, version 2.2 for further guidelines, see [Info](#)

**Enter up to 10 sentences with up to 10 words each and press the Analyze button.
Always use a verb and a subject in your sentences**

Feedback Language български ▾

```
mannen smile  
et gutt ikke kommer  
jeg ser gutten sin  
Ola skammer
```

Analyze

generate

Думата "smile" е в инфинитив, но трябва да бъде в минало или в сегашно време. [More description](#)

generate

Детерминаторът трябва да има същия род, число и определеност като на съществителното име, което модифицира. [More description](#)
В главно изречение не е възможно наречие да предхожда личния глагол.

generate

Притежателното рефлексивно местоимение "sin" не съвпада по лице с думата, към която се отнася. Вместо него използвайте "min". [More description](#)

generate

Глаголът "skammer" изисква възвратен обект. [More description](#)

Norwegian Grammar Tutor

Demo with ACE, version 2.2 for further guidelines, see [Info](#)

**Enter up to 10 sentences with up to 10 words each and press the Analyze button.
Always use a verb and a subject in your sentences**

Feedback Language Arabic ▼

```
mannen smile  
et gutt ikke kommer  
jeg ser gutten sin  
Ola skammer
```

Analyze

generate

في صيغة الفعل المصدرية ولكن يجوز أن تكون في صيغة الحاضر أم الماضي "smile" كلمة [More description](#)

generate

A determiner must have the same gender, number and definiteness as the noun it modifies. [More description](#)
في عبارة أساسية لا يمكن إن الظرف يسبق الفعل الأساسي

generate

The reflexive pronoun "sin" does not match the person of the word it refers back to. Try using "min" instead. [More description](#)

generate

يحتاج فعل "skammer" إلى ضمير انعكاسي كمفعول به. [More description](#)

Controllo della grammatica norvegese

Incatenata a [Messaggi di feedback](#).

Questo sistema è additato a procurare feedback grammatico alle frasi che hai scelto.

Per favore, cliccare sull'immagine del troll  per trovare il controllo grammatico.

Nella finestra che appare, scrivere una o più frasi (fino a dieci frasi alla volta, cambiando riga dopo ogni frase, e ogni frase costituendo non più di 10 parole) e premere il pulsante 'Analizza'. Se la frase è grammaticalmente ben formata, avrai la risposta:

```
Una frase norvegese ben formata.
```

Se la frase non è gramaticalmente ben formata, ci sono due possibilità:

Diagnosi dell'errore

La prima possibilità è che il sistema ti dice che cosa nella frase è sbagliato. Ad esempio, alla frase sgrammaticata

```
Jeg liker du.
```

dove la forma di soggetto "du" è usata come oggetto, avrai la risposta:

```
La parola "du" è marcata con il caso sbagliato, prova invece a usare "deg".
```

Insieme con un tale messaggio di errore, può apparire un pulsante 'Generate'. Se lo clicchi avrai una nuova finestra mostrando una versione raccomandata della frase desiderata, in questo caso:

```
Jeg liker deg.
```

Aspects of the technical infrastructure

The following slides summarize the main aspects of the webservices.

Webservices for Parsing and Generating

- The webservices are used in our web applications
- The services use the AceScriptLibrary to parse and generate with ACE
- The library reads from and writes to Unix processes. We have one process for parsing and another for generating.
- Download library from: <http://regdili.hf.ntnu.no:8081/norsourceDownload/>

Parameters for Parse Webservice

- statement, the statement to be parsed
- syntax, true/false in order to receive syntax trees or not
- readings, in the interval [1..10]
- mrs, true/false in order to receive MRSEs or not
- client, the name of the calling application for our sentence log

Parameters for Generate Webservice

- mrs, the MRS in LKB format to be generated
- statement, for our statement log
- client, the name of the calling application for our sentence log

http://regdili.hf.ntnu.no:8081/bongram/rest/parse?statement=mannen+smiler

```
<?xml version="1.0" encoding="UTF-8" standalone="true"?>
- <parse>
  <readings>1</readings>
  <error source="unknown">NOTE: 1 readings, added 310 / 31 edges to chart (6 fully instantiated, 26 actives used, 4 passives used) RAM: 1616k </error>
  - <syntax-tree top="n1">
    <terminal parent="n4" name="mannen" id="n5"/>
    + <node parent="n3" name="mann_masc-reganim" id="n4" num="3" end="1" beg="0" pct="0.000000">
    + <node parent="n2" name="sg_masc-def-noun-lxm-lrule" id="n3" num="461" end="1" beg="0" pct="-0.146200">
    + <node parent="n1" name="sg_def_m_final-full_irule" id="n2" num="462" end="1" beg="0" pct="-0.146200">
    <terminal parent="n7" name="smiler" id="n8"/>
    - <node parent="n6" name="smile_intr" id="n7" num="5" end="2" beg="1" pct="0.000000">
      <argument id="n8" number="1"/>
    </node>
    + <node parent="n1" name="pres-infl_rule" id="n6" num="463" end="2" beg="1" pct="-0.260702">
    - <node name="head-subject-rule" id="n1" num="464" end="2" beg="0" pct="-0.433458">
      <argument id="n2" number="1"/>
      <argument id="n6" number="2"/>
    </node>
  </syntax-tree>
  - <mrs index="e1" ltop="h0">
    + <ep end="-1" beg="-1" quote="true" pred="_mann_n_rel" label="h3">
    + <ep end="-1" beg="-1" quote="true" pred="_def_q_rel" label="h4">
    - <ep end="-1" beg="-1" quote="true" pred="_smile_v-intr_rel" label="h7">
      <arg name="ARG0" ref="e1"/>
      <arg name="ARG1" ref="x2"/>
    </ep>
    <feature name="SORT" value="verb-act-specification" key="e1"/>
    <feature name="SF" value="prop" key="e1"/>
    <feature name="E.TENSE" value="pres" key="e1"/>
    <feature name="E.MOOD" value="indicative" key="e1"/>
    <feature name="E.ASPECT" value="semsort" key="e1"/>
    <feature name="WH" value="-" key="x2"/>
    <feature name="BOUNDED" value="+" key="x2"/>
    <feature name="PNG.NG.NUM" value="sing" key="x2"/>
    <feature name="PNG.NG.GEN" value="m" key="x2"/>
    <feature name="PNG.PERS" value="thirdpers" key="x2"/>
    <hcons label="h3" hole="h5" rel="qeq"/>
  </mrs>
  <status>ok</status>
</parse>
```

Forward

Ever improving the grammar, internally, as well as regarding robustness-enabling facilities (see slides from last year's presentation).

Getting users for the scrabble and the sparrer, but also for the API as such. Thus, we are exploring the possibility of driving traffic to the API as *parsing as a service* for interested third parties.

And for use of the LKB architecture more generally in semantic and typological research, resuming work on *Grammar Induction from IGT* (cf. Hellan and Beermann 2014).

Some references

Hellan, L., Tore Bruland, Elias Aamot, Mads H. Sandøy (2013): A Grammar Sparrer for Norwegian. *NoDaLiDa 2013. NEALT Proceedings Series, volume 16.*

Hellan, L., D. Beermann, T. Bruland, M.E.K. Dakubu, and M. Marimon (2014) MultiVal: Towards a multilingual valence lexicon. *LREC 2014.*

Hellan, L. and D. Beermann (2011, 2014). Inducing grammars from IGT. Z. Vetulani and J. Mariani (eds.) *Human Language Technologies as a Challenge for Computer Science and Linguistics.* Springer.