Some applications around a Deep Grammar

Lars Hellan, Dorothee Beermann, Tore Bruland,

Tormod Haugland, Elias Aamot

Presented at LTC 2015 and 2017, Poznan

The cluster

Chronologically first in the development were two lexical repositories, *TROLL* in the late 80ies and *NorKompLex* in the late 90ies, the latter partly extending the former. They were followed by a computational grammar built on the LKB platform (cf. Copestake 2002) using HPSG (cf. Pollard and Sag 1994), called *NorSource*, started in 2001 and still being developed, with information from the lexical repositories as its main 'start capital'. NorSource in turn has the following offsprings: an on-line language learning tool called the *Norwegian Grammar Sparrer* running on NorSource (from 2011 on); a large multi-lingual online valency lexicon, *MultiVal*, in its construction development based crucially on NorSource (from 2013 on), a POS-tagger constructed from the information in NorSource (2014), and a valence corpus - *Norwegian Valency Corpus*.

From our perspective, NorSource may be seen as the architectural center point of these applications, with a typed feature structure (TFS) build-up which accommodates all the information in the lexical repositories, and with a computational TFS-based processing system which allows this information to be operative both in the general parser and in the further applications.

NorSource (‘Norwegian HPSG Resource Grammar')

As a so-called *Deep Computational Grammar*, NorSource sustains a *generic* parser (not restricted with regard to style of text or domain of use) representing wide lexical coverage, encoding linguistically well motivated morpho-syntactic and semantic analyses of nearly all aspects of the grammar, and applying this knowledge in the parsing process such that every parse reflects this knowledge.

NorSource was started in 2001 in the EU-project *DeepThought*, and is still being maintained and developed, conducted at NTNU. It has been sponsored by EU, NFR, NTNU. Online access, for description: http://typecraft.org/tc2wiki/Norwegian_HPSG_grammar_NorSource .

Webdemo: http://regdili.hf.ntnu.no:8081/linguisticAce/parse

The NorSource code files are downloadable from: http://www.nb.no/sprakbanken/show?serial=sbr-32&lang=en

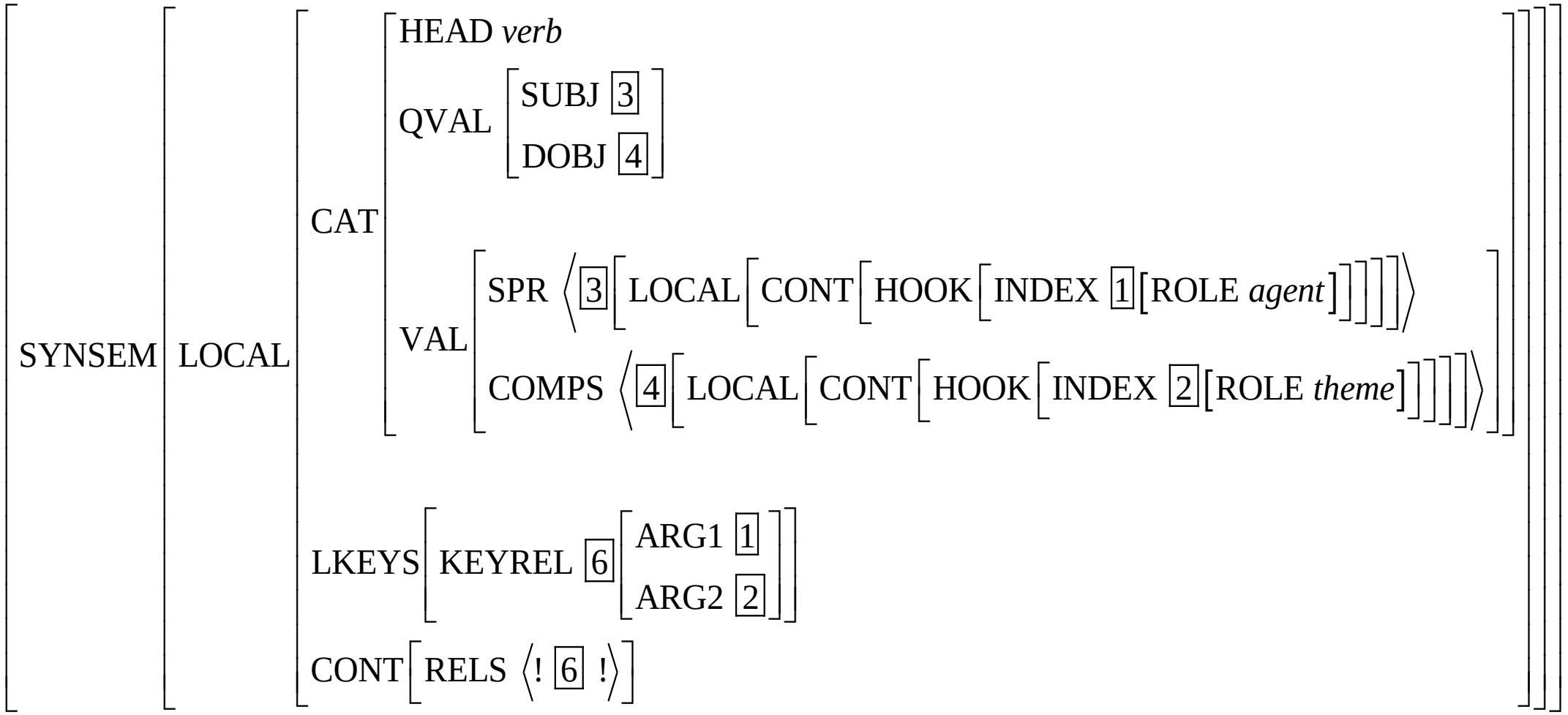
The system LKB as such can be downloaded from http://moin.delph-in.net/LkbTop.

NorSource

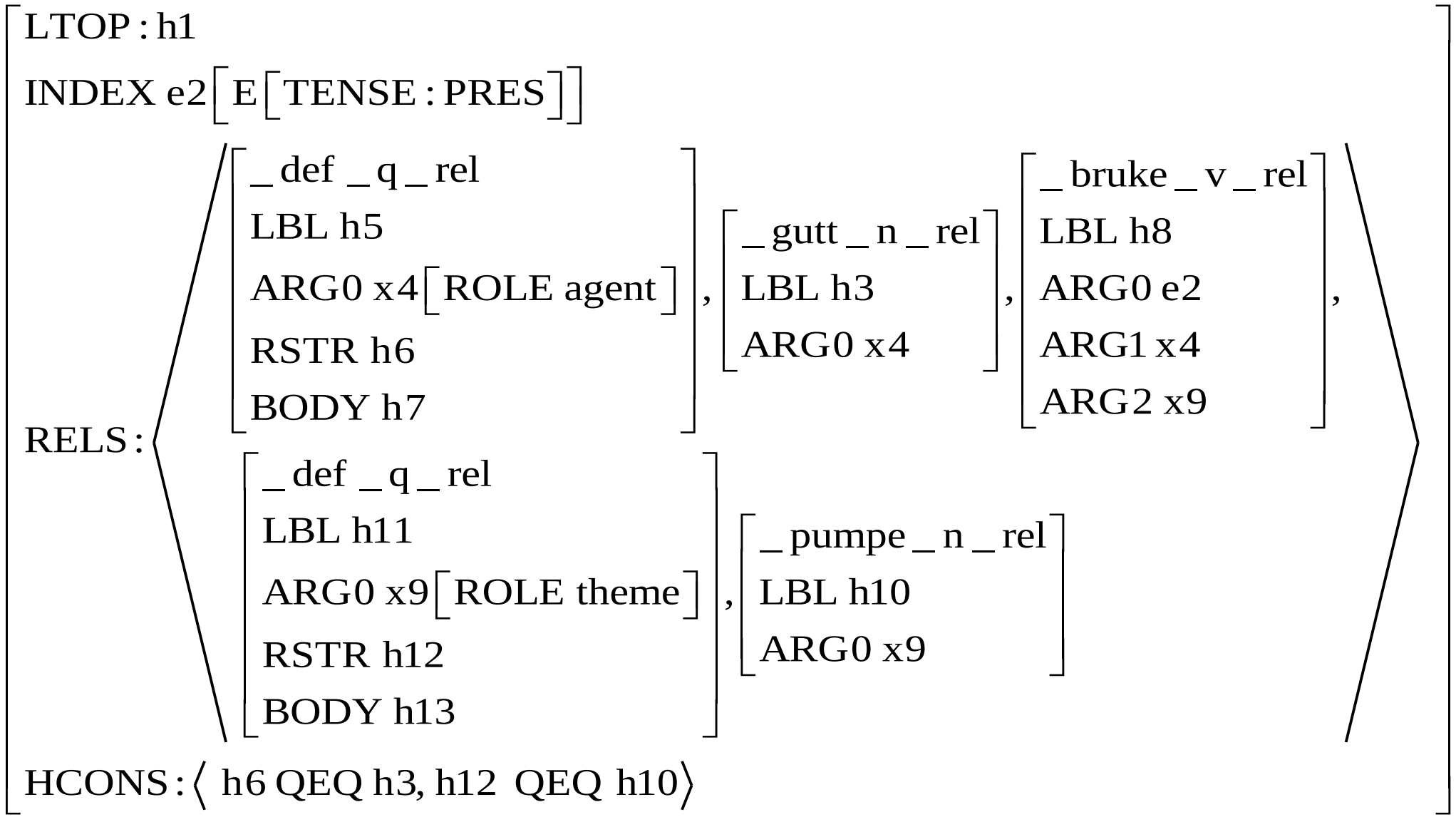
NorSource has as its formal and theoretical framework *Head-Driven Phrase Structure Grammar* (*HPSG*) (Pollard and Sag 1994, Sag et al. 2003), on which the computational project initiative *LinGO* at CSLI, Stanford, was started, using the *LKB platform* (Copestake 2002), which is a general platform with the format of typed feature-structures (TFS), and has integrated in it a format of semantic representation called *Minimal Recursion Semantics* (‘MRS’; cf. Copestake et al. 2005).

Before year 2000 there were three grammars in this framework, viz. the *English Resource Grammar* ('ERG'), the Japanese grammar 'Jacy', and the German grammar 'GG'. Essential to the development of further grammars of this type was the *HPSG Grammar Matrix* (‘the *Matrix*’; see Bender et al. 2002, 2010), which was mainly based on ERG, and had its first phase of deployment during the EU-project *DeepThought* (2002-4). NorSource was the first grammar based on this platform, and the since then growing family of grammars (by now 10-12 well developed grammars) is now hosted by the DELPH-IN consortium. http://moin.delph-in.net/

Grammatical representation of the type v-tr-suAg_obTh



MRS representation for ***Gutten bruker pumpen*** *'the boy uses the pump'*



NorSource - stages

- 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 *Interoperability* phase (2010-).
- **Phase 1** resided in the building of a basic core grammar around the Matrix skeleton (using the Matrix versions 0.1 – 0.6, as they developed; this included the MRS system). This stage included the accommodation of a 80,000 entries lexicon imported from the previously established resources TROLL and NorKompLex, where a verb valence code and a code for inflectional paradigms constituted major parts. Main publications from this period are: Hellan and Haugereid 2002, Hellan 2003.
- **Phase 2** resided in the development of a fine-grained ontology and computing system of spatial and temporal relations, amenable to grammatical systems across languages and typologies, and a detailed semantics of comparative constructions. The grammar was also used as a part of a small Norwegian-Japanese MT system. In this period, the inflectional system was thoroughly revised. Main publications: Hellan and Beermann (2004), Beermann et al. (2004), Beermann and Hellan (2005), Hellan and Beermann (2012).

  This phase features a tdl-file with the semantics of spatial and temporal relations for prepositions: http://typecraft.org/tc2wiki/Norwegian_HPSG_grammar_NorSource, which can be used across all the Matrix grammars.

NorSource – stages (2)

Phase 3 was devoted to a thorough revision of the valence code, to accommodate a cross-linguistically defined classification system of valence and construction types. Main publications : Hellan (2008), Hellan and Dakubu (2010), Dakubu and Hellan forthcoming. Opens also for *Grammar Induction*.

Phase 4 has resided in the development of applications:

- A 'Grammar Sparrer', as described in Hellan et al. 2013, accessed at A Norwegian Grammar Sparrer. This is a construct along the lines of Bender et al. 2004, and Suppes et al. 2014, falling within the overall initiatives described in Heift and Schultze 2007, where specific types of grammatical mistakes are accommodated by 'mal-rules' in an extended 'mal'-version of the grammar, and parses involving such mal-phenomena are reported to the user as tutoring instructions. This system has been running as a webdemo since 2011.
- A Multilingual Valence repository, called ***MultiVal***, based on NorSource and three further LKB grammars: The Spanish Resource Grammar, the Bulgarian grammar BURGER, and a grammar of Ga. See slides below. http://regdili.hf.ntnu.no:8081/multilanguage_valence_demo/multivalence
- An initial version of a POS-tagger of Norwegian, reflecting the lexical inventory of the grammar, which amounts to appx. 85000 lexical entries, and a large number of proper names of various categories. The tagger currently offers all available POS-alternatives for a given word. See web access at http://regdili.hf.ntnu.no:8081/webtagger/tagger.
- An automated procedure for generating a valence corpus of Norwegian, the corpus situated and searchable in TypeCraft. https://typecraft.org/tc2wiki/Norwegian_Valency_Corpus

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

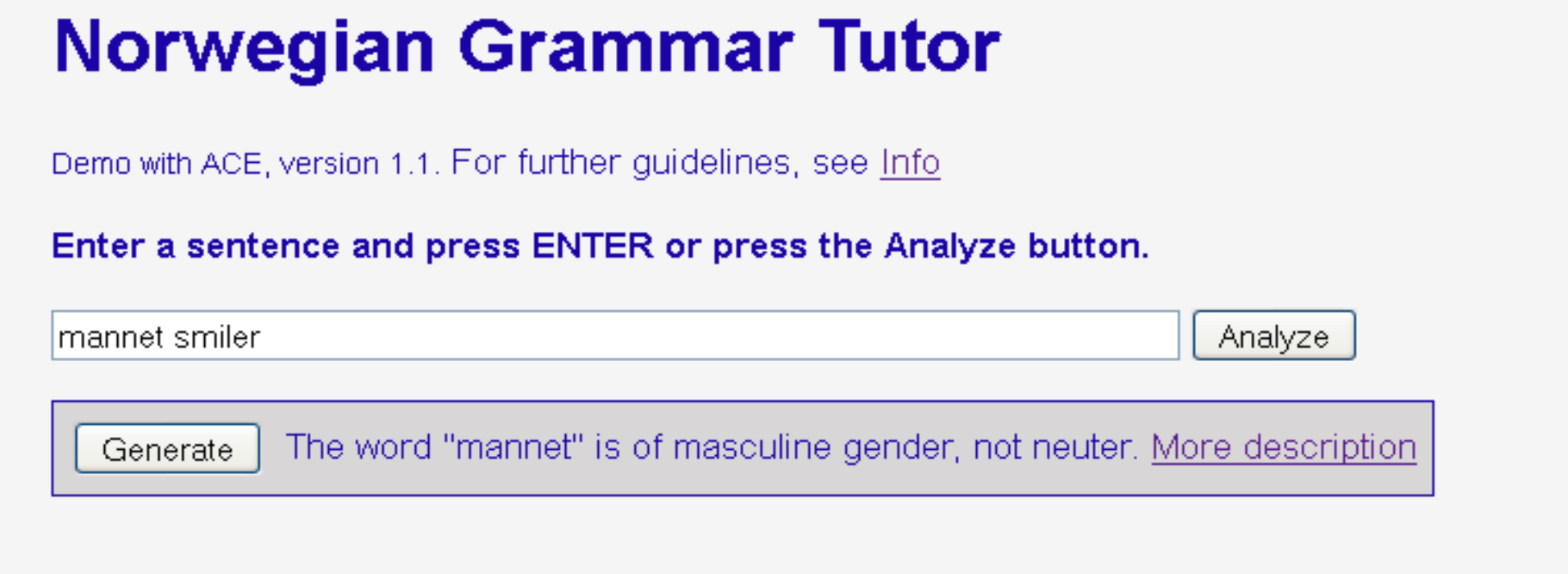
http://typecraft.org/tc2wiki/A_Norwegian_Grammar_Sparrer

An introduction to its ‘mal-grammar’-based design is given in Hellan et al. 2013.

Its basics, as developed in 2011-2013, are indicated on the following two slides:

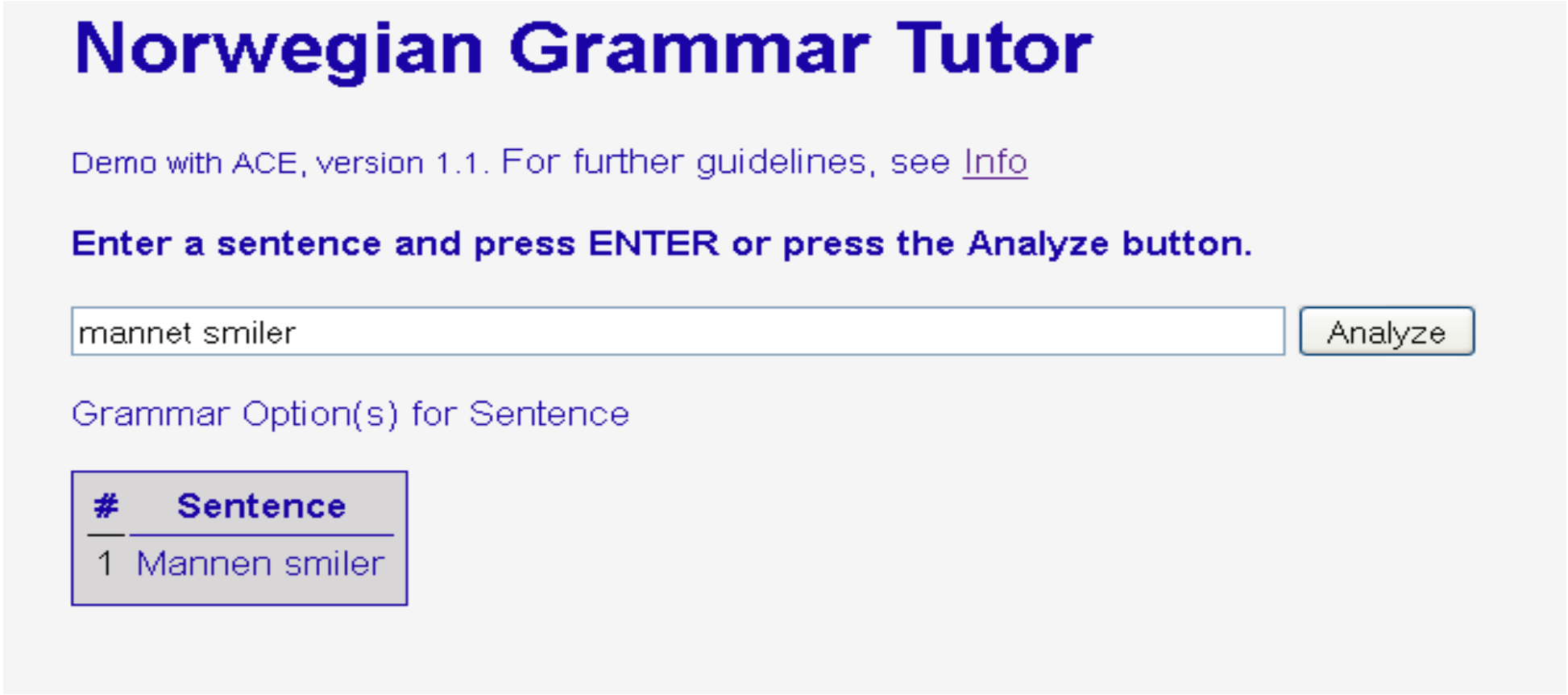
The procedure - 1

- If you enter an ungrammatical sentence …
  - you receive an error message (see lowest line underneath).
  - If the MRS constructed for the sentence by the ’mal-grammar’ is wellformed, a button for ’Generate’ appears (see below), by which a ’correct’ version of the sentence can be evoked.



The procedure – 2:

Using ’Generate’ to see an acceptable option



During the last year

The interface now accepts ***batches of up to 10 freely chosen sentences***, each with max. 10 words

Responses are given not only in *English*, but also in *Polish, Italian, German, Bulgarian, Chinese, Norwegian,* and partly *Arabic.*

The design with *freely chosen inputs* requires a large grammar and lexicon – 84 000 entries.

The number of actions (processing a batch of sentences, or doing a generation) has been around 300 per day during 2017 and 2018.

A corpus of input strings is being accumulated. (But we keep no track of users.)

The system now sits in a virtual server at the faculty of Humanities, NTNU.

The system functions on a Creative Commons license basis.

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.

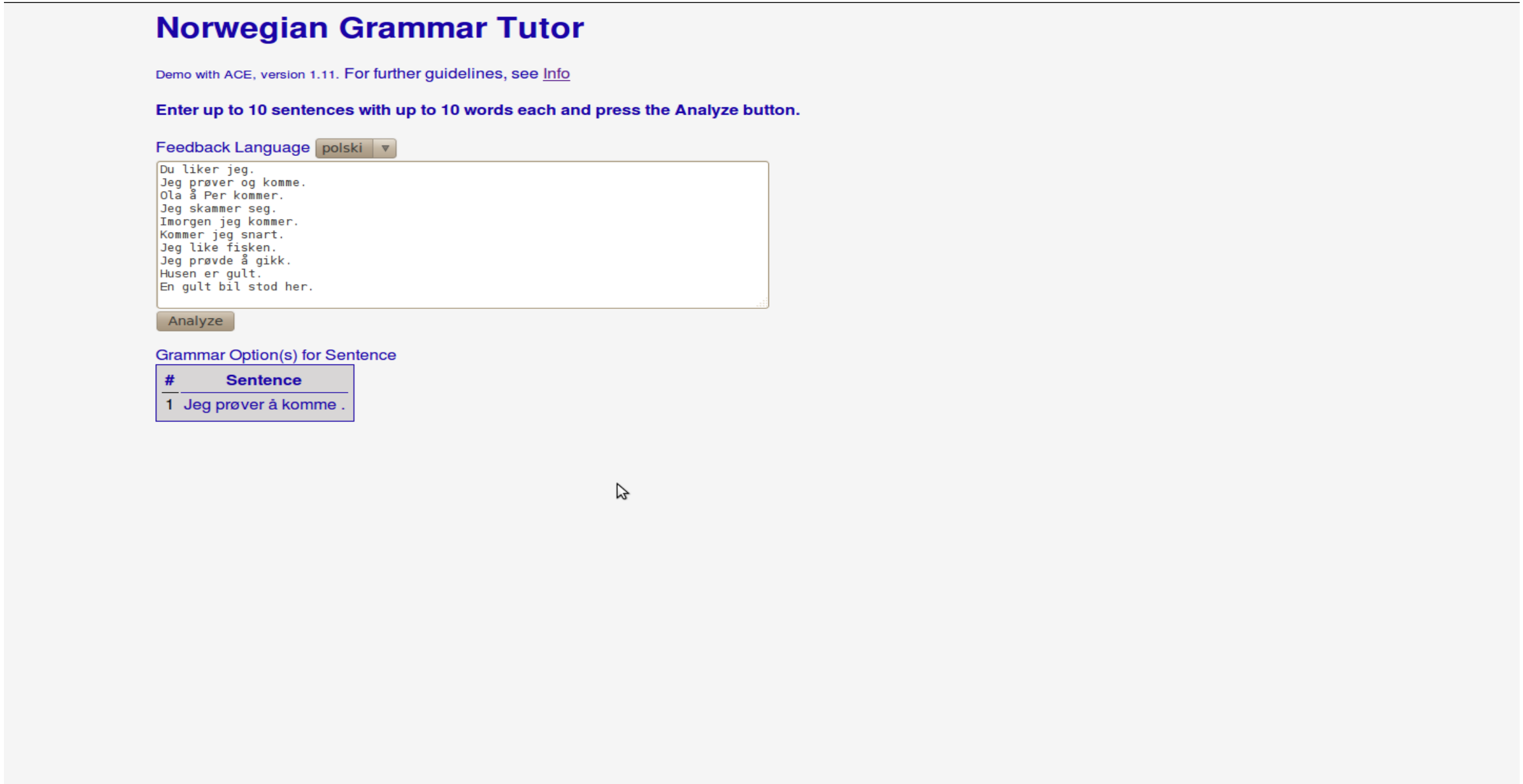
Feedback Language polski

```
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 któego 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 for ***sentence 2***



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

The system by now hosts 4 languages, with altogether 40 000 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 imported from 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 of the system per February 2014 (before Bulgarian got added), see Hellan et al., LREC 2014.

The following slide shows search results for ‘intransitive’, for verb starting with “s”.

The subsequent slide in turn shows information as it looks for a given verb, and shows two features of interoperability with other applications – *TypeCraft* and *ImagAct*.

Version 1.4 (for further guidelines, see Info)

Languages:

☑ Norwegian ☑ Ga ☑ Spanish ☑ Bulgarian

Search fields:

Verb lexeme: s

Syntactic Arguments:

Function: intransitive

Situation:

Aspect:

Type:

Search | Count | Clear | Download

Ga	show	sa	NP
Ga	show	sa	NP
Norwegian Bokmål	show	sabbe_intr-dir	NP
Norwegian Bokmål	show	sabbe_iv	NP
Norwegian Bokmål	show	safe_iv	NP
Norwegian Bokmål	show	safte_iv	NP
Norwegian Bokmål	show	sage_iv	NP
Norwegian Bokmål	show	sake_iv	NP
Norwegian Bokmål	show	sakke_iv	NP
Norwegian Bokmål	show	sakse_iv	NP
Norwegian Bokmål	show	saktne_iv	NP
Spanish	show	salivar_v	NP
Norwegian Bokmål	show	saluttere_iv	NP
Bulgarian	show	salyutiram_v1	+
Norwegian Bokmål	show	samarbeide_iv	NP
Norwegian Bokmål	show	sameksistere_iv	NP
Bulgarian	show	samoobladavam_v1	+

For the button SHOW:

Automatic import of urls for glossed examples from ***TypeCraft*** has been defined, and links to ***ImagAct*** scene videos are being added – here for *Marit tar seg på kinnet:*

Language	Norwegian Bokmål
Verb id	ta_tr-detachposs-refl
Syntactic Arguments	NP+NPrefl+PP
FCT	transReflxWithOblique
SIT	ternaryPossessorDetachment
Aspect	
Verb type	v-trObl-obRefl_oblPRTOFob
Example of type	Ola klør seg på ryggen
Orthography	ta
English gloss	[take] – only through TypeCraft link
Example	[Marit tar seg på kinnet] – only through TypeCraft link
Free translation	[Mary touches her cheek] – only through TypeCraft link
TypeCraft URL	http://typecraft.org/tc2/ntceditor.html#2790,45468
ImagAct URL	http://www.imagact.it/imagact/sceneMetadata.seam?sceneId=54&cid=9995

Transfer of information from the ’provenance’ grammars into MultiVal

The information encoded in a verb type is unfolded through a conversion script, exemplified below with one out of the nearly 300 rewrite rules. The leftmost item in this rule is a lexical type, which reflects both grammatical and semantic properties. This rule rewrites the type symbol ‘v-ditr’ (‘ditransitive headed by verb’), into the syntactic argument structure (SAS) counterpart ‘NP+NP+NP’, the functional specification ‘ditransitive’, and the semantic specification of a three-place relation.

v-ditr => SAS: “NP+NP+NP”

FCT: ditrans

SIT: ternaryRel

This information is available in the online interface, whereby exactly the amount of consolidated information available in the other members of the cluster is now available also in an online query interface.

The Valence Corpus

Valence corpora are most often been built manually, or by statistical methods where hand annotation plays a crucial role.

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)

In some cases valence corpora, possibly in conjunction with tree-banks, are used in the construction of computational grammars.

Osenova (2011); Patujek and Przepiórkowski (2016)

Here we go the opposite way, exporting information from the deep grammar to an IGT corpus, whereby sentences in the corpus serve as categorized examples of the verb valence types as defined in the grammar.

The corpus and its information

- The corpus consists of 22000 sentences imported from the Leipzig Corpus Collection, all with the standard TypeCraft IGT annotation and with valence information for each verb occurrence, given in the form exemplified above. The valence information is stated relative to the ACTIVE form of the verb, even if the example provided is in passive. When doing search you can use either of these types of labels. The codes are explained and exemplified as follows:
- 'ConstructionLabel' at Verbconstructions cross-linguistically – Introduction, Valence Profile Norwegian, Valence Profile English.
- SAS at Valency label 'SAS'
- FCT at Valency label 'FCT'
- Joint illustrations of them all are given in Valency code illustrations.

How to search

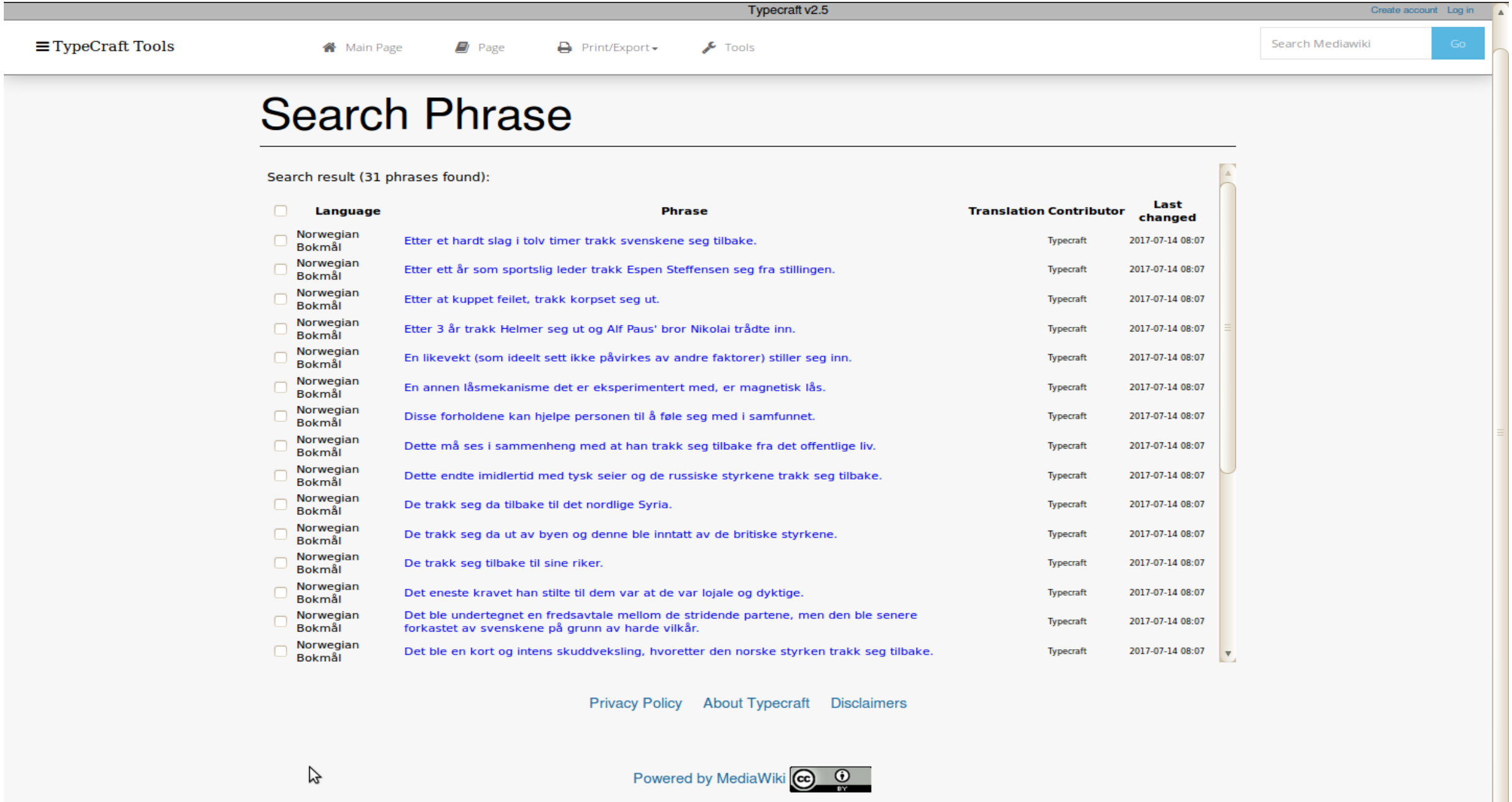
You can search relative to valence type in general, or specifically for a given verb, where the verb can be stated by citation form or by its actually occurring form. The search interface is the standard one for TypeCraft:

`TypeCraft Tools` (in upper left corner) `-> TypeCraft Search -> Phrase search`.

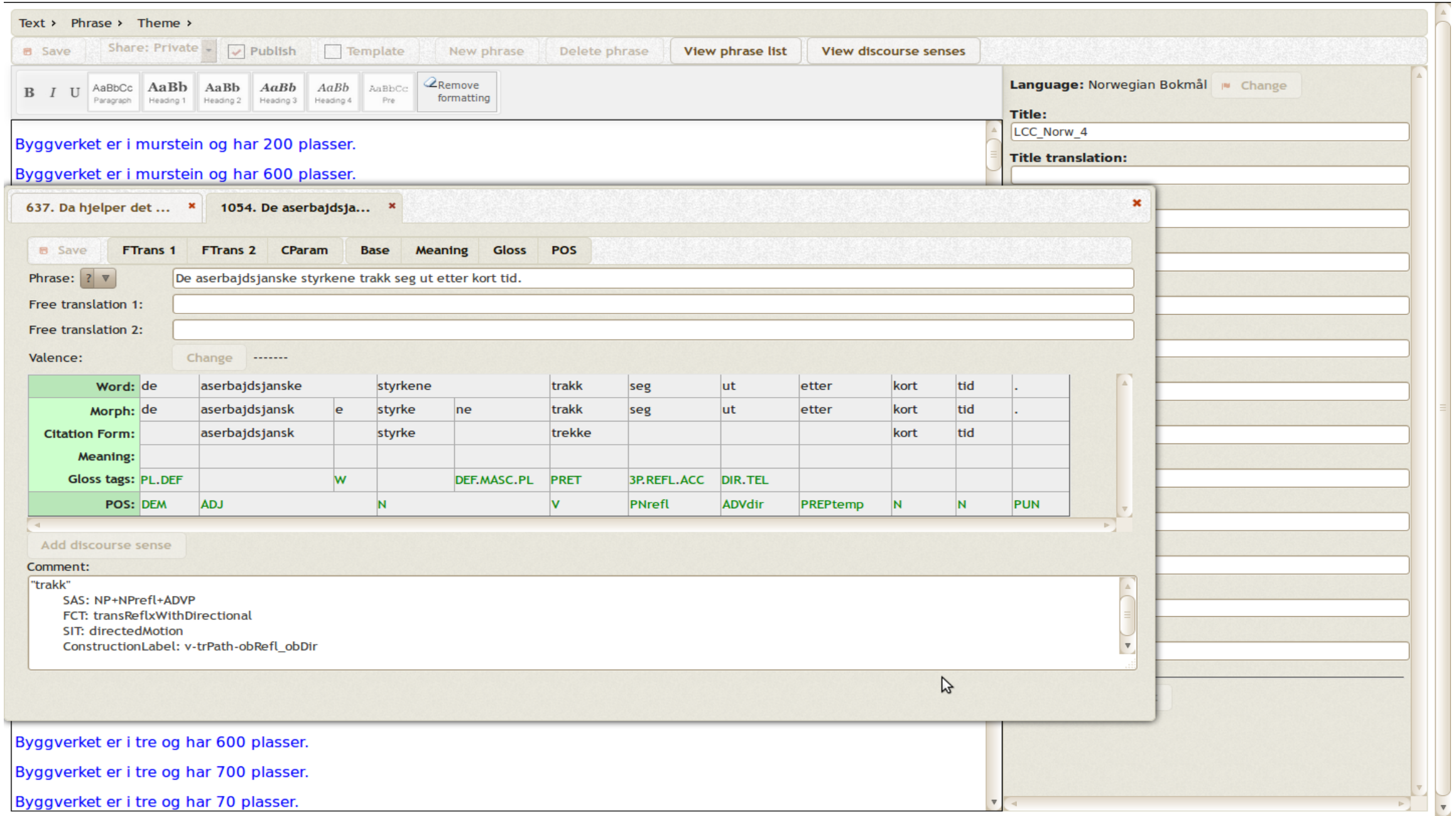
On this page choose 'Norwegian Bokmål' from the Language menu; at 'Phrase level', write (or glue) the valence label into the slot 'Phrase description'. If you want to search also relative to verb, enter the exact form of the verb under 'Word level - Exact form'. (The slot for its citation form is 'Morpheme level - Exact base form', however this search option is temporarily disabled. The same holds for any other search for morphological properties when done in conjunction with 'Phrase description'.)

A verb lexicon with valence types given in the ConstructionLabel format is given in Valence lexicon.

Search result for the frame type 'reflexive + directional'



From hyperlink to instance



Illustrating Valence + IGT 'normal form'

String: Jeg vet at hun forbauset ordføreren

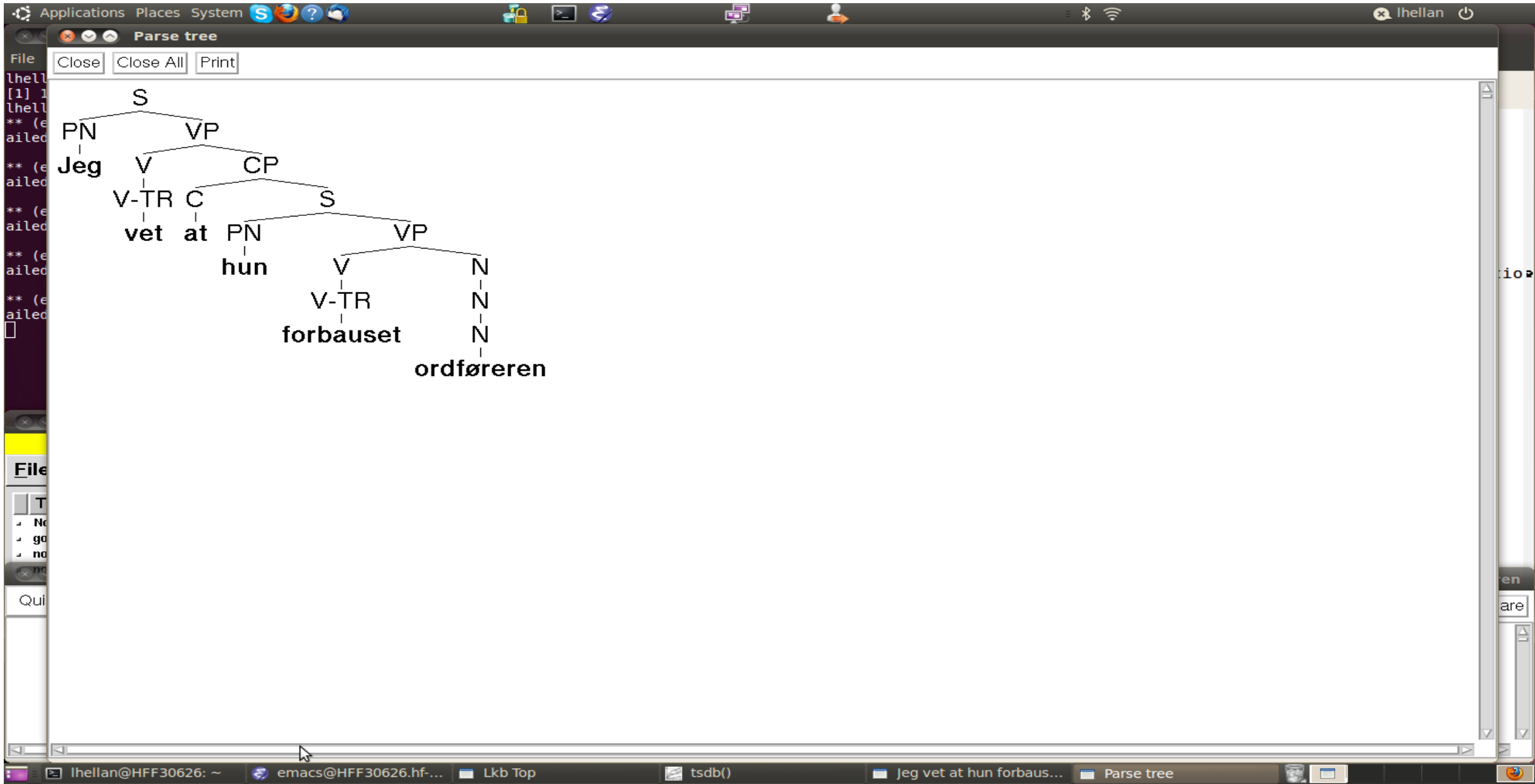
Free translation: I know that she surprised the mayor

Morph	Jeg	|	vet	|	at	|	hun	|	forbause	|	t	|	ordfører	|	en
Cit.		|	vite	|		|		|	forbause			|	ordfører		
Gloss	|1.SG.NOM	|	PRES	|	DECL	|	3.SG.FEM	|		|	PRET			|	DEF.SG.MASC
POS	|PN	|	V	|	COMP	|	PN	|	V				N		

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

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

One view of what the grammar produces



Alternative view of the parse tree

head-subject-rule
 jeg_perspron
  jeg
 head-verb-inf-or-s-comp-rule
  pres-infl_rule
   vite_subord_vlxm
    vet
  head-complementizer-comp-fin-rule
   at_subord
    at
   head-subject-rule
    hun_perspron
     hun
    head-verb-comp-rule
     pret-nonfstr-et_infl_rule
      forbause_tv_vlxm
       forbauset
     sg_def_m_final-full_irule
      sg-masc-def-noun-lxm-lrule
       ordfører_n_masc_nlxm
        ordføreren

Using the parse tree for valence extraction

We assign a valence value to every verb occurrence in a sentence. For *vet* a look-up in the verb lexemes file establishes that the identifier in question carries the type *v-tr-obDECL* (cf. the simplified view of a verb entry in (a)), and look-up in a file establishing correspondences between the CL code and the SAS and FCT codes yields (b).

a. vite_subord_vlxm := v-tr-obDECL

b. v-tr-obDECL =>

SAS: "NP+Sdecl";

FCT: transWithSentCompl

From these correspondences the following part of the Figure is established:

vet: SAS: NP+Sdecl

FCT: transWithSentCompl

ConstructionLabel: v-tr-obDECL

The files in which these look-ups are made count 12,000 entries corresponding to (a), and about 400 conversions corresponding to (b).

Using the parse tree for POS and GLOSS extraction

Inflectional processing is done via ‘rules’, stated in a form exemplified below (for verbs 22 such rules, for nouns 28, and for adjectives 38).

pret-nonfstr-et_infl_rule :=
%suffix (e a) (e et) (es es) (es edes)
infl-pret-verb-word &
[ARGS <[INFLECTION nonfstr-et]>].

This rule is mentioned in the tree for *forbauset*, reflected in the lines

pret-nonfstr-et_infl_rule
forbause_tv_vlxm
forbauset

stating that the form *forbauset* has been derived from the lemma form *forbause* by the application of this rule.

Using the parse tree for POS and GLOSS extraction

The appropriate GLOSS tag in TC will be PRET, and this is assigned through the mapping rule below to the GLOSS line in TC:

pret-nonfstr_infl_rule = PRET

There are altogether 75 mapping rules from Norsource inflection rules to TC GLOSS tags. Most of them apply simply to rule names, more examples for verbs are given below

- a. ppart-finalstr-dd_infl_rule = PRF
- b. s-passive_s_infl_rule = PRES.PASS
- c. pl_def_m-or-f_light-e_irule = PL.DEF
- d. sg_def_n_light-e_irule = SG.DEF.NEUT

Using the parse tree for POS and GLOSS extraction

GLOSS for constant words

- a. fordi_comp = CAUS
- b. idet_prep-time = TEMP
- c. mer_cmpar-reg = CMPR
- d. seg_refl = 3P.REFL.ACC
- e. en_indef-art = SG.MASC.INDEF

POS for words according to entry suffix, or to word as a whole (constituting most of the 472 mappings to POS tags :

- a. nlxm = N
- b. alxm = ADJ
- c. vlxm = V
- d. dirtel-end-p = PREPdir
- e. reg-p-loc = PREPplc
- f. mer_cmpar-mass = QUANT

Using the parse tree for POS and GLOSS extraction

GLOSS for constant words

- a. fordi_comp = CAUS
- b. idet_prep-time = TEMP
- c. mer_cmpar-reg = CMPR
- d. seg_refl = 3P.REFL.ACC
- e. en_indef-art = SG.MASC.INDEF

POS for words according to entry suffix, or to word as a whole (constituting most of the 472 mappings to POS tags :

- a. nlxm = N
- b. alxm = ADJ
- c. vlxm = V
- d. dirtel-end-p = PREPdir
- e. reg-p-loc = PREPplc
- f. mer_cmpar-mass = QUANT

Assessment 1

- It is the first time that a full grammar has been mapped for GLOSS and POS to TC, and for TC this is thus an interesting situation of testing its inventory. Essentially all of the GLOSS and POS tags required for representing features and word classes in Norwegian are represented in TC, so that there are very few cases where a tag has to be created for this specific mapping.
- It is in turn a benefit for Norsource that its features can be displayed in the fashion provided by TC, its GLOSS-type features otherwise being barely interpretable from the grammar-native feature structures coming with a syntactic parse.

Assessment 2

- The quality of the valence information depends on the quality of the deep parser, that is, a deep parser combines syntactic and semantic parsing with the recognition of predicate-argument structure, and our valence corpus therefore will be only as good as the parser is in handling these grammatical dependencies. Moreover the quality of the corpus depends on the conversion itself which is not without complexity, as has been indicated, so that mistakes could arise, and, per the automatic design, 'infect' a large number of sentences.
- Yet an obvious advantage of the method is that, once analyses are deemed plausible, one can in relatively little time obtain a comprehensive valence corpus. The Norwegian IGT-valence corpus is now in its trial phase and we expect some feedback to the parser which arises from the pairing of IGT and valence code.

Assessment 3

- The present corpus is strictly a valence corpus, combined with standard IGT, and does not aim at including semantic information beyond what pertains to argument structure. Thereby essential parts of the semantic analysis machinery of a grammar of the present kind is being ignored (such as the module Minimal Recursion Semantics (cf. Copestake et al. 2005), which is an integral part of Norsource); but could in principle be incorporated at later stages.
- A valence resource should also include a valence lexicon, where each verb is specified for all the frames in which it can occur, preferably with access to selected examples from a corpus, and to a link to a larger corpus and its seach interface, where for instance also frequency data based on the corpus can be called upon. A lexeme based overview of valence frames is available.

Conclusion - general

No application in the overall cluster was purportedly designed with a view to supporting the other applications (except that the lexicon applications perhaps might have a parser as a possible employment), thus each one was created in its own right.

None of them were computationally innovative, but rather based on solid techniques and platforms.

The linguistic content was also solid and as ‘deep’ as any computational application can have it, but not theoretically innovative per se.

In all parts, the applications can be easily understood by linguists and computational linguists, a circumstance which has allowed for a certain change of maintainers over time, and which makes the further sustainability and development of the resources a realistic prospect.

Conclusion – *Klart språk*

It will seem that a *Klart språk* corpus can be processed by the same procedures as here described. The key challenge will be to read from the parse results the types of factors that are significant to ’Klart språk’ (assuming that the grammar identifies them in the first place), and represent them in the annotated version of the corpus in a readily searchable fashion.

It is also conceivable that at ’Clarity-checker’, in the same spirit as the Grammar Checker, could be developed, once the grammar identifies the relevant factors, and one has the means to develop a suitable interface.

References

Bender, E. M., Drellishak, S., Fokkens, A., Poulson, L. and Saleem, S. (2010). Grammar Customization. In *Research on Language & Computation*, Volume 8, Number 1, 23-72.

Bruland, T. (2011). Creating TypeGram data from TypeCraft. Presentation at *India 2011*, NTNU.

Bruland, T. (2013). Building World Event Representations From Linguistic Representations. PhD dissertation, NTNU.

Copestake, A. (2002). *Implementing Typed Feature Structure Grammars*. CSLI Publications.

Copestake, A., D. Flickinger, I. Sag and C. Pollard. (2005). Minimal Recursion Semantics: an Introduction. *Journal of Research on Language and Computation*. 281-332.

Heift, T., and M. Schulze. (2007). *Errors and Intelligence in Computer-Assisted Language Learning: Parsers and Pedagogues*. Routledge, New York.

Hellan, L. (2008) From Grammar-Independent Construction Enumeration to Lexical Types in Computational Grammars. COLING (http://www.aclweb.org/anthology-new/W/W08/#1700).

Hellan, L., Johnsen, L. and Pitz, A. TROLL. Ms, University of Trondheim. 1989

Hellan, L. and D. Beermann. 2005. Classification of Prepositional Senses for Deep Grammar Applications. In: Kordoni, V. and A. Villavicencio (eds.) *Proceedings of the 2nd ACL-Sigsem Workshop on The Linguistic Dimensions of Prepositions and their Use in Computational Linguistics Formalisms and Applications,* Colchester, UK.

References contd

Hellan, L. and Dakubu, M.E.K. (2010). *Identifying Verb Constructions Cross-linguistically*. SLAVOB series 6.3, Univ. of Ghana.

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

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

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

Oepen, S. and Carroll, J. (2000). Performance profiling for parser engineering. *Natural Language Engineering*, *6. Issue on Efficient Processing with HPSG*, 81 – 97.

Pollard, C. and Sag, I. (1994). *Head-Driven Phrase Structure Grammar.* Chicago University Press.

Sag, I., Wasow, T. and Bender, E. *Syntactic Theory*. A formal introduction. Stanford: CSLI Publications.

Suppes, P, T. Liang, E.E. Macken and D. Flickinger (2014) “Positive technological and negative pre-test-score effects in a four-year assessment of low socioeconomic status K-8 student learning in computer-based Math and Language Arts courses ", Computers & Education, 71, pp. 23-32.

Tesnière, L. (1959). *Éleménts de syntaxe structurale*. Paris: Klincksieck.

Further references

- Copestake, Ann, Dan Flickinger, Ivan Sag and Carl Pollard. 2005. Minimal Recursion Semantics: An Introduction. *Journal of Research on Language and Computation*. 281-332.
- 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 Verb Valency*. Amsterdam & Philadelphia: John Benjamins Publishing Co.
- Goldhahn, D., Eckart, T., Quasthoff, U. (2012). Building Large Monolingual Dictionaries at the Leipzig Corpora Collection: From 100 to 200 Languages. In: Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC'12), 2012.
- Adam Przepiórkowski, El´zbieta Hajnicz, Agnieszka Patejuk, Marcin Woli´nski, Filip Skwarski, and Marek ´Swidzi´nski. (2014). Walenty: Towards a comprehensive valence dictionary of Polish. In Calzolari et al. (eds) 2014.
- Nicoletta Calzolari, Khalid Choukri, Thierry Declerck, Hrafn Loftsson, Bente Maegaard, Joseph Mariani, Asuncion Moreno, Jan Odijk, and Stelios Piperidis, (eds.) Proceedings of the 9th International Conference on Language Resources and Evaluation (LREC 2014), pages 2785–2792, Reykjavík, Iceland. ELRA.
- Osenova, Petya (2011). Localizing a Core HPSG-based Grammar for Bulgarian. In: Hanna Hedeland, Thomas Schmidt, Kai Worner (eds.) *Multilingual Resources and Multilingual Applications*, *Proceedings of GSCL 2011*, ISSN 0176-599X, Hamburg, pp. 175-180.
- Patejuk, Agnieszka (2016). Integrating a rich external valency dictionary with an implemented XLE/LFG grammar. In Doug Arnold, Miriam Butt, Berthold Crysmann, Tracy Holloway King, Stefan Müller (eds.) *Proceedings of the Joint 2016 Conference on Head-driven Phrase Structure Grammar and Lexical Functional Grammar.* Stanford: CSLI Publications, pp 520-540.
- Pollard, Carl and Ivan A. Sag (1994). *Head-Driven Phrase Structure Grammar.* Chicago: Chicago University Press.
- Sag, Ivan A., Thomas Wasow and Emily Bender. 2003. *Syntactic Theory*. CSLI Publications, Stanford.