

Lars Hellan, NTNU

May 18, 2015

TROLL ('Trondheim Linguistic Lexicon') 1987-91

Conducted at: Linguistics Department, NTNU (then: University of Trondheim)

Sponsored by: NFR (The Norwegian Research Council, then: Norges Allmenntvitskapsfelag forskningsråd (NAVF)) and University of Trondheim.

Project responsible: Lars Hellan. Further participants: Lars G. Johnsen, Margaret Magnus, Anneliese Pitz, Tor Åfarli, Hanneke van Hoof, Elisabeth Wennevold Silva.

Main results:

Developing a format for lexical description suitable for a computational dictionary, in the areas of syntax, semantics and morphology, applying it to descriptions of the verbal systems of Norwegian, German and Dutch.

Main publication: Hellan et al. 1989.

TROLL has about 10 000 lemmas of nouns, verbs and adjectives, acquired by the project. Its main focus is verbs, hosting around 2000 verb entries, with detailed information about valency, thus constituting a *valency lexicon* of Norwegian. This information includes both individual frames of the lemmas, and information about the array of valency alternation frames relative to each frame. It antedates the design of Levin 1993, and what has later become the design associated with *VerbNet*. Each verb entry thus provides

- (i) a valency frame of the verb, called its *basic frame*, and
- (ii) an index of the set of alternating frames.

The combination of (i) and (ii) is called the verb's *template*. The set of alternating frames corresponds to what in *VerbNet* is called the *verb class* of the verb. Two formats of indices are used: numerical, and brief formulaic *macros* summarizing the content.

In the definition of each basic frame, each argument is defined by semantic role, syntactic category and grammatical function, constituting a triple <role, category, function> for each argument ('rkf-triple').

For each template, the alternation possibilities are stated as *derivational* rules (D-rules), which can apply in ordered sequences, a bit like transformations in a transformational grammar, being defined over the information represented in the argument triples. Unlike the frame alternations defined in *VerbNet* and systems akin to it, D-rules in *TROLL* apply not only within the category of verbs ('intra-lex D-rules'), but also across categories, as when adjectives are derived from verbs, and nouns in turn from adjectives. With the 'basic' frames and the possible results of applications of sequences of D-rules, altogether 155 possible frames are defined as such by the system, and the total number of frame occurrences across the 1000 'basic' verb entries lies around 4000.

The following is an example of a verb entry

```
(1)
101707 :=
[para : v2(real <- ascii),cat : [c : "V",infl : "inf"],
 real : [ascii : "bruke"],morph : "[bruk([],[e])]",
 templ : (6, 27)].
```

101707 is the entry identifier; v2 states the inflectional class of the verb – for v2 the distinguishing property is the past tense form -te); 'real <- ascii' states that the spelling given is in ascii; 'cat' introduces a statement about part of speech of the entry and of which inflectional form of the verb is entered in the line below, namely as "bruke". The verb is mono-morphemic. The valency templates associated with the verb carry the indices '6' and '27'. Information about the verb's argument properties is thus represented only by these indices in the entry, the indices themselves being defined elsewhere.

An entry for a multi-morphemic verb is given below, with the stem morph preceding the non-stem morphs:

(2)

```
101681 :=
[para : v2(real <- ascii),cat : [c : "V",infl : "inf"],
real : [ascii : "beviser"],morph : "[vis([be],[e])]",
templ : (6, 25, 24, m311 + m322)].
```

Here the fourth template given is 'm311 + m322', which is a codified representation of a specific combination of arguments, the numbers corresponding to specific roles, categories, or functions, according to where they occur in the ordering. Such encoding is used when the argument combination is too specific to warrant a general template definition.

In the folder **Troll_files**, the following files are provided summarizing the specification codes and categories used, and definitions of derivational rules and sequences, all relative to verbs:

```
rkf_and_template-macro_codes
D-rules_for_verbs_intralex
D-rules_crosslexeme
norsuff
template_macros
template_numbering
templates_with_intralex_derivseq
templates_with_cross_lexeme_derivseq
templates_with_derivseq_simplified
```

Less formal descriptions of the derivational rules and of the basic templates are provided in

```
Basic templates
D-rules
```

together with a more extensive introduction to the system:

```
Introduction
```

These will all be provided in a future edition of this description.

Examples of noun entries and adjective entries are given in (3) and (4), designed like verb entries except for nouns lacking template specifications:

(3)

```
101948 :=
[para : a1(real <- ascii),cat : [c : "A",infl : "neu"],
real : [ascii : "barnerik"],
morph : "[barn([],[e]),rik([],[])]",templ : m911].
```

(4)

```
101949 :=
[para : n1(real <- ascii),cat : [c : "N",infl : "sing"],
real : [ascii : "barnerim"],
morph : "[barne([],[]),rim([],[])]",templ : ""].
```

A specification of inflectional codes for all three parts of speech is given in

```
Inflection codes
```

also supplied in a future edition of this report. The full lexicon files of TROLL are provided in **Troll_Lexfiler.tar.gz**.

TROLL was partly carried further in **NordLex (1992-94)**, a Nordic cooperation project sponsored by NOS-H with partners at NTNU (Univ. of Trondheim), Copenhagen Business School, Lunds Universitet, and University of Iceland, Reykjavik. Main coordinator was Lars Hellan, and participants at Copenhagen Business School: Finn Sørensen (site manager), Irene Byron, Michael Herslund; at Lunds Universitet: Christer Platzack (site manager), Lena Ekberg, Gunlög Josefsson; at University of Iceland: Eiríkur Rögnvaldsson (site manager), Adalsteinn Eythorsson; at University of Trondheim: Ingebjørg Tonne, Anja Seibert, Jon Atle Gulla, Sjur Moshagen. This project applied relevant formats of TROLL to the other Scandinavian languages, and developed especially aspects of semantic analysis. As a 'bud' project of TROLL was developed a 'Sign' framework of linguistic analysis; a formal integration of the two was suggested in Gulla's 'Sign Expansion Theory' ('SET'); the 'Sign Model' as such developed many similarities with HPSG (Pollard and Sag 1994), which allowed for a formal integration with HPSG from 1998 on, in part through the computational grammar Norsource (see below), although with many aspects of semantic analysis having to be explored further.

Another subsequent project to TROLL was:

NordKompLex (NKL) 1996-2001

Conducted at NTNU

Sponsored by: NFR, Telenor, NTNU

Project responsible: Torbjørn Nordgård

Participants: Bodil Aurstad, Eli Sætherø Andenes, Jardar Eggesbø Abrahamsen, Bente Moxness, Anne Nylund, Kristin Eide, Arne Kjell Foldvik, Jørn Almborg

Main results:

- Providing large scale computational dictionaries of both bokmål and nynorsk, covering inflectional properties for all word classes, argument structure descriptions for verbs, and phonological specifications for all entries (base forms and inflected forms, as a full form lexicon).

Main publication: Nordgård 1998.

In this project, about 100 of the 150 types accounted for in TROLL were selected in a 'flat' type system, with argument structure specifications done as in TROLL, using essentially the 'rkf' format, but not distinguishing between basic and derived types, and thus with no derivational structure defined.¹ With many lexemes thus appearing in multiple entries, and with a large import of verbs from *Bokmålsordboka*, the verb part of the NorKomplex dictionary counts around 10 000 verb entries; adding adjectives and nouns, NorKompLex hosts around 80 000 lemmas.

The NorKompLex dictionary has been a resource both in commercial and academic applications. It is an essential basis for *Norsk Ordbank* which has an online search interface.² It has also served in the build-up of the lexicon of two computational grammars, NorSource, see below, and the LFG grammar *NorGram*.³

(A presentation of NorKompLex will soon be available in Språkbanken.)

Norsource 2001 – still in operation

Conducted at NTNU

Sponsored by: EU 5th framework, NTNU, NFR, all in small amounts, and through own resources throughout.

¹ Of course keeping a process like 'passive' still as a potentiality feature. The verb type labels here have the form of 'macros' such as intrans1, intrans2, intrans3,

² <http://www.edd.uio.no/perl/search/search.cgi?appid=72&tabid=1106>

³ <http://clarino.uib.no/iness/xle-web>

Project responsible: Lars Hellan

Participants: Dorothee Beermann, Petter Haugereid, Ben Waldron, Tore Bruland, Elias Aamot, Mads Hustad

See the presentation in Språkbanken.

This grammar, based on the LKB platform for HPSG computational grammars, bases its main lexicon on NorKompLex, but has a smaller lexicon based on the TROLL lexicon, which was converted to LKB format in 2006 by Lars Johnsen. ‘Productive’ alternation rules like Passive are encoded in Norsource as Lexical Rules, in keeping with the HPSG design, but many of the alternation and derivational relationships encoded in TROLL are still represented only in a ‘flat’ format, for concerns of efficiency of parsing.

Some of the code alignments between the TROLL format and the Norsource format are found in the folder *from TROLL to Norsource*, and the entries in (5) show the Norsource encoding corresponding to (1)-(4). It will be noted that (apart from the regular ‘lexical rule’ Passive) one of the derivational processes encoded in TROLL, namely ‘Object deletion’, is reflected in the verb entries (5a,b); a challenging further step of grammar development will be to reflect more of the D-rules from TROLL in Norsource, including the cross-lexeme relations.

(5)

a.

```
bruke_tv-reg := trans-arg1-2-verb-lexeme &
  [STEM <"bruke">,
  INFLECTION nonfstr,
  MORPH-SPLIT <"=bruk=e">,
  SYNSEM.LKEYS.KEYREL.PRED "_bruke_v_rel",
  SYNSEM.LOCAL.CONT.HOOK.XARG.ROLE initiator,
  SYNSEM.LOCAL.CONT.HOOK.VARG.ROLE non-initiator,
  SYNSEM.DERIV-POT.CAN-OBJ-DEL -].
```

b.

```
bevis_tv-reg := trans-arg1-2-verb-lexeme &
  [STEM <"bevis">,
  INFLECTION nonfstr,
  MORPH-SPLIT <"=be=vis=e">,
  SYNSEM.LKEYS.KEYREL.PRED "_bevis_v_rel",
  SYNSEM.LOCAL.CONT.HOOK.XARG.ROLE initiator,
  SYNSEM.LOCAL.CONT.HOOK.VARG.ROLE non-initiator,
  SYNSEM.DERIV-POT.CAN-OBJ-DEL -].
```

c.

```
barnerik_stnd-adj := stnd-adj-lxm &
  [STEM <"barnerik">,
  INFLECTION nonfstr-ic,
  MORPH-SPLIT <"=barn=e","=rik=">,
  SYNSEM.LOCAL.CAT.HEAD.KEYS.KEY adj-k,
  SYNSEM.LKEYS.KEYREL.PRED "_barnerik_j_rel"].
```

d.

```
barnerim_neut := neut-noun-lxm &
  [STEM <"barnerim">,
  INFLECTION final-full-nshort,
  MORPH-SPLIT <"=barne=","=rim=">,
  SYNSEM.LOCAL.CAT.HEAD.KEYS.KEY nom-k,
  SYNSEM.LKEYS.KEYREL.PRED "_barnerim_n_rel"].
```

The counterparts of the full lexical files of TROLL in the Norsource version are supplied in *Troll-based Norsource lex-files.tar*.

Two further developments of the TROLL initiative can be mentioned.

(i) A further elaboration of the macro system for templates, called *Construction Labeling*; see Hellan 2008, Hellan and Dakubu 2010.

(ii) An online search interface into the valency information encoded in Norsource, i.e. an online *valency bank*; it is combined with information from similar grammars of Bulgarian, Ga and Spanish; see Hellan et al. 2014.

Thus, in the latter application, a search facility is provided for lexical information, and in Norsource, a computational parser using lexical information, thereby accomplishing functionalities envisaged in TROLL, although not implementable at the time of TROLL. Although the implemented systems go beyond the content of TROLL in most respects, with respect to frame alternation and cross-lexeme relationships, the information in TROLL still remains to be implemented in the applications mentioned and any corresponding application.

References

- Copestake, A. (2002). *Implementing Typed Feature Structure Grammars*. CSLI Publications, Stanford.
- Hellan 2008: From Grammar-Independent Construction Enumeration to Lexical Types in Computational Grammars. Paper presented at COLING, Workshop on Grammar Engineering Across Frameworks (GEAF) Manchester, August 2008 (<http://www.aclweb.org/anthology-new/W/W08/#1700>).
- Hellan, L., L. Johnsen and A. Pitz. 1989. TROLL. Ms., Univ. of Trondheim.
- 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 Dakubu, M.E.K. (2010). *Identifying Verb Constructions Cross-linguistically*. SLAVOB series 6.3, Univ. of Ghana (http://www.typecraft.org/w/images/d/db/1_Introlabels_SLAVOB-final.pdf).
- Levin, B. 1993. *English Verb Classes and Alternations*. University of Chicago Press, Chicago, IL.
- Nordgård, T.(1998) "Norwegian Computational Lexicon (NorKompLeks)". Proceedings of the 11th Nordic Conference of Computational Linguistics NODALIDA 98. CST, Copenhagen.