Trondheim 2020

Running applications, valence, construction typology, grammar of Ga, ‘primitives’ of argument structure

Lars Hellan

DELPH-IN Summit, July 2020

Contents

1. Applications running (references on slide 15)

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

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

https://typecraft.org/tc2wiki/TypeGram

(Norsource technical development and maintenance by Tore Bruland, functions hosted by TypeCraft maintained by Dorothee Beermann, Pavel Mihaylov, Tormod Haugland.)

2. TFS-based construction ontology

Typed Feature Structures (TFS) as used in LKB are well suited for building construction ontologies, offering the complexity required for representing syntax and semantics of sentential constructions. Labels for aspects of constructions in terms of which construction hierarchies can be modeled, were presented in Hellan and Dakubu 2010, Dakubu and hellan 2017; argument structure and situational semantics are the features exhibited in most detail.

References:

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 Publ. 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.

A model of grammar reflecting these compositional aspects was developed around 2010, providing a construction type system covering Germanic, Kwa and Bantu, under the name '***tg***' for 'typological grammar'. It has a feature system based on the early ESSLLI training grammars (cf. Copestake 2002), thus formally simpler than that of the standard Delph-In grammars; its type-files are found at https://typecraft.org/tc2wiki/TypeGram. A goal is that the parse result of a sentence perspicuously displays the construction type realized by the sentence. The integration of the **tg** system as part of a grammar design is described in (Hellan 2019), and further in (Hellan 2020a,b) (on b - see section 5).

References:

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

Hellan, Lars. 2020a. Interoperable Semantic Annotation. LREC workshop ISA-16, 6th Joint ACL-ISO Workshop on Interoperable Semantic Annotation | workshop web site | Proceedings (pdf)

3. Developing Norwegian valence resources

The Norwegian resource is called ***NorVal***. It is partly derived from the Norsource lexicon.

By a '*lexval*' we mean a verb entry formally specified relative to one valence frame only. NorVal at the moment has 12,845 lexvals . We show

3.1 Number of verbs with one frame only (3,228 verbs)

3.2 Number of polyvalent verbs (3026 verbs, with entries distributed over 9,837 lexvals).

3.3 Relevance to valence representation in a grammar lexicon:

3.1 3,228 verbs with a unique frame; most frequent:

- 2,145 for transitive
- 656 for intransitive
- 88 for transitive with light reflexive object
- 65 for intransitive with oblique
- 42 for transitive with a particle
- 36 for intransitive with a directional subject
- 28 for ditransitive
- 22 for impersonal
- 16 for transitive with directional object, 15 for transitive plus oblique, 14 for ditransitive with light reflexive as indirect object , 12 for transitive with light reflexive object plus oblique , 12 for transitive with light reflexive directional object, 11 for transitive with a particle, 10 for transitive with a particle and light reflexive object, …

3 verbs have a subject-controlled infinitive as unique frame: *nedlate seg til å*. *plikte å*, *unnlate å*

No unique frame consists of a declarative or interrogative complement, an extraposed clause, or an absolute infinitive.

3.2. 3026 polyvalent verbs

A *valence pod* is the set of lexvals defining a polyvalent verb, here illustrated for the verb :***“be”*** *(‘ask’, ‘request’,’beg’, ‘pray’)* with 11 members:

be:

{__intr & -for__intrObl-oblN & -om__intrObl-oblEqSuInf & -om__intrObl-oblN & __tr & -om__trObl-oblEqObInf & -om__trObl-oblEqSuInf & -om__trObl-oblN & -på__trObl-oblN & -til__intrObl-oblN & __ditr-obEqIobBareinf }

Meaning of specification codes, see next slide, and for illustrating examples, the slide thereafter.

Meaning of specifications:

__intr - intransitive

-for__intrObl-oblN - intransitive with an oblique complement headed by the preposition “for”

-om__intrObl-oblEqSuInf - intransitive with an oblique complement headed by the preposition “om”, and where the preposition governs an infinitive equi-controlled by the subject

-om__intrObl-oblN - intransitive with an oblique complement headed by the preposition “om”

__tr – transitive

-om__trObl-oblEqObInf - transitive with an oblique complement headed by the preposition “om”, and where the preposition governs an infinitive equi-controlled by the object

-om__trObl-oblEqSuInf - transitive with an oblique complement headed by the preposition “om”, and where the preposition governs an infinitive equi-controlled by the subject

-om__trObl-oblN - transitive with an oblique complement headed by the preposition “om”

-på__trObl-oblN - transitive with an oblique complement headed by the preposition “på”

-til__intrObl-oblN - transitive with an oblique complement headed by the preposition “til”

__ditr-obEqIobBareinf – ditransitive where the object is a bare infinitive equi-controlled by the indirect object

Examples

be__intr & vi ber ‘we pray’,

be__tr & vi ber dem ‘we ask/invite them’,

be__ditr-obEqIobBareinf & hun ber dem komme ‘she asks them to come’,

be-for__intrObl-oblN & vi ber for dem ‘we pray for them’,

be-om__intrObl-oblN & vi ber om tillatelse ‘we ask for permission’,

be-til__intrObl-oblN & vi ber til gud ‘we pray to god’,

be-om__trObl-oblEqObInf & vi ber dem om å reise seg ‘we ask them to stand up’,

be-om__trObl-oblEqSuInf & jeg ber dem om å få sitte ‘we ask them to be permitted to sit’,

be-om__intrObl-oblEqSuInf & hun ber om å få komme ‘she asks to be permitted to come’,

be-om__trObl-oblN & de ber dem om lov ‘they ask them for permission’,

be-på__trObl-oblN & vi ber dem på middag ‘we invite them on supper’.

Polyvalence pods

What one can do with pod-assemblies:

- Find correlations pod-structures and meaning, intuitively perceived – if enough correlations, that can serve as an ‘extensional’ clue to verb meaning. (
- By comparing pods, detect omissions in existing specifications or new possibilities in valence patterns.
- Both of these can be extended to cross-linguistic studies.

3.3 Relevance to valence specification in a grammar lexicon

A lexicon like the one here envisaged can be used in an HPSG grammar (and was in this case derived from such a lexicon); there are at least two points where either structure could make do with one lexval where the other has two.

1. Optionality: Possibility of reducing a pod, or elements in a pod, to a frame with optional elements.
2. Underspecification: Possibility of reducing alternative linear sequences to just one frame.

We illustrate the points.

Possibility of reducing (part of) a pod to a frame with optional elements.

A standard move in lkb-grammars is to, whenever a verb has both a transitive and an intransitive frame, to define just one frame, where the object is marked as optional. Where the grammar has an explicit semantics, that will normally mean assigning just one semantic representation, whereby the second argument is treated as an ‘implicit argument’ relative to the intransitive frame. Similarly for other alternations.

The NorVal lexicon, just like the lexicon in the Norsource grammar, makes no use of optionality specifications; but the NorVal lexicon offers a possibility of investigating pods at a large scale to reveal to what extent optionality marking, applied as mentioned regarding semantics, may be a sound option.

Underspecification of frames relative to alternative linear sequences

In NorVal transitive particle constructions such as “spise opp maten” (‘eat up the food’) vs. “spise maten opp” (‘eat the food up’) have a common frame marked as ‘*trPrtcl*’ simply, whereas in an explicit parsing grammar (like Norsource) a valence frame must state the linear order explicitly. There are 762 verbs for which such a pair of specifications would be needed.

A possible underspecification strategy could be to define frames (a) and (b), with the frame label ‘trPrtcl’ corresponding to the partial specification (c), which in parsing could match either (a) or (b) (‘GF’ for ‘grammatical function’ , ‘ARG2’ for the object and ‘ARG’3 for the particle in the semantic representation).

(a) … GF [OBJ #1 & [LOCAL … INDEX #3], PRTCL #2 & [LOCAL … INDEX #4]],

… COMPS <#1, #2> , … […, ARG2 #3, ARG3 #4], …

(a) … GF [OBJ #1 & [LOCAL … INDEX #3], PRTCL #2 & [LOCAL … INDEX #4]],

… COMPS <#2, #1 > , … […, ARG2 #3, ARG3 #4], …

(c) trPrtcl := … GF [OBJ #1 & [LOCAL … INDEX #3], PRTCL #2 & [LOCAL … INDEX #4]],

… […, ARG2 #3, ARG3 #4]

This remains to be tried. (Could be relevant also for languages with ‘scrambling’.)

4. Developing comparative valence resources

This research (Lars Hellan and Dorothee Beermann) addresses Norwegian, based on NorVal (cf. section 3), and German, with partial build-up of resources. The first area in focus is infinitival constructions across the languages. (In addition to colleagues mentioned in the presentation titles below, also Melanie Siegel takes part.)

Preliminary presentations:

Dorothee Beermann, Lars Hellan, Tormod Haugland, Anna Struck. 2019. Et korpusbasert sammenlignende valensprosjekt for norsk og tysk. MONS 2019, DUGNADSGRUPPE: Korpusbasert grammatikkforsking.

Uwe Quasthoff, Lars Hellan, Erik Körner, Thomas Eckart, Dirk Goldhahn, Dorothee Beermann. 2020. Typical Sentences as a Resource for Valence. LREC 2020. http://www.lrec-conf.org/proceedings/lrec2020/index.html

Lars Hellan. 2019. Verb valence lexicons, and comparing them. DELPH-IN Summit, 2019, Cambridge.

5. Ga computational grammar

The RAIL workshop at LREC this year offered a possibility for presenting a Ga computational grammar, embedded in the construction typological TFS system mentioned in section 2 and thus implemented through **tg** (cf. section 2); cf. Hellan 2020b.

While the **tg** design is easy to process for the human eye, grammars based on the system lack many of the features of Matrix-grammars, and a recent step is integrating the **tg** design in a full-fledged Matrix-type grammar of Ga. This grammar extends the grammar presented at HPSG2007 (Dakubu et al. 2007)), with Matrix-style increased coverage of serial verb constructions and three-argument constructions.

The grammar can be downloaded from

https://typecraft.org/tc2wiki/Ga_computational_HPSG_grammar.

'Primitives' of Argument Structure

Among features one can explore through this grammar are what one may call 'primitives of argument structure', that is, analytic concepts needed to derive correct syntax and semantics in the domain of argument structure. For constructions like serial verb constructions and NP-embedded dependencies (discussed at some length in Hellan 2020b), and likewise, relative to a Norwegian grammar, for some of the preposition-governed infinitival constructions illustrated on slides 9 and 10, relevant 'primitives' used in the Ga grammar and Norsource are valence lists, grammatical functions, status as 'ARG1', 'ARG2' … within 'relation', and status as 'XARG' and similar constructs inside of HOOK projections. Also used in the literature is 'ARG-ST', and explicit reference to semantic roles, although not in these grammars. As an exercise in 'constructive empiricism' it can be worthwile to try out various combinations of such concepts in explicit grammars, matched against linguistically motivated analyses of given phenomena.

References concerning Ga computational grammar:

Dakubu, M.E.K., L. Hellan and D. Beermann. 2007. Verb Sequencing Constraints in Ga: Serial Verb Constructions and the Extended Verb Complex. In St. Müller (ed) *Proceedings of the 14th International Conference on Head-Driven Phrase Structure Grammar.* Stanford: CSLI Publications. (/http://csli-publications.stanford.edu/)

Lars Hellan. 2020b. 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

Further background references

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