'Klart språk' in Norwegian – concept and goals from a linguistic viewpoint

Lars Hellan

NTNU, 07.05.2018

A 'Verwaltungs'-initiative

The Norwegian Government has encouraged public institutions to optimise the way in which their communication with 'the public' can get understood.

In all countries, 'the public' is divided into those who can in principle intellectually take care of themselves, and those who cannot.

Countries differ in how much help and protection is given to the latter group. In Germany, *Aktion Mensch* and *Leichte Sprache* are examples of such attention being given. In Norway, there are counterparts to these initiatives.

The purpose of *Klart språk* is thus to make the interaction between 'Verwaltung' and the intellectually full-functional part of the population as efficient as possible.

A premiss is that the State is benevolent: if a citizen doesn't understand a message from the Verwaltung, he or she risks a lot of trouble, and the goal of *Klart språk* is to reduce such risk.

What *is* **'Klart språk' -** 'transparent language' – Beermann 2017

Language Dimension	
Precision	Language is about things and situations it refers to, yet precision is not always transparent neither referentially, structurally nor semantically, but rather ambiguous on all levels. Precision relates to the effectiveness to disambiguate.
Complexity	Complexity depends on the words used, their frequency, and distribution across registers. Important factors influencing complexity are the relationship between predicates and arguments, length of utterances and the depth of structural embedding. Important factors for the complexity of discourse are its rhetorical structure and the density of indexical and scopal expressions.
Adequacy	Adequacy reflects the extent to which speakers recognize an utterance as part of their language, and understand the essence of the communication without instructions or training.

Our focus in *Klart språk*

In *Klart språk* **Precision** is mandatory and non-negotiable (in *Leichte Sprache* perhaps less so).

We can perhaps study **Adequacy**, in terms of what people actually get out of texts, and come up with proposals for strategies to be used in *brukerundersøkelser*, for those institutions who want to follow up their Klarspråk initiative with seeing how intended improvements actually work. But for them, that will probably be a long shot, and thus also for us.

Thus, our focus must be **Complexity**.

That will hang nicely together with linguistic analysis in general. And in our setting, comparative/contrastive analysis. (Since we will probably be most 'useful' if we do stuff that we like and are good at, all the better.)

Activity with *Klart språk*

Many public institutions put work into *Klart språk*.

Mattilsynet, Lotteristiftelsen, Skatteetaten, Språkrådet, …

The *Juridiske fakultet* in Oslo holds a fairly big Scandinavian conference on the topic later this month.

What must be ensured to be 'clear' of course depends on the matters being communicated, and so it is good that the institutions themselves do the main work. But what is the work?

Partly developing guidelines for authors of letters and documents of various types. Partly engaging in the authoring of the documents directly.

Both Mattilsynet and Lotteristiftelsen are willing to share with us how they work in these respects.

But who is then 'we'? A small group of linguists. What is our role – giving tipps about what is 'clear' language? Hardly. Do research into what 'clear' language really is? Or develop automatic detection tools for less clear language?

Research into what characterizes 'clear' language

(1) Constructing corpora of 'clear', resp. 'not clear', language.

(2) Annotating these texts relative to factors suspected/hypothesized to influence degree of clarity (for instance, 'compound noun', 'light verb construction', 'passive', are candidate factors/features).

(3) Doing statistics over the texts relative to what we have annotated.

Texts to use for (1) we hope to get from the institutions mentioned.

The factors influencing clarity – we should have a well worked out set of features and criteria for how to apply them before we start annotating.

Given that we know beforehand which texts are clear and which not, it is only after (3) that we know whether these factors really reflect clarity.

If the features do not yield a clear distinction between the texts, then one must try with other features.

The methodology is at least transparent. The queezy aspect is the reliance on what people have declared as 'klart' and 'ikke klart'. Are the criteria clear?

Automated procedures

Given our financial resources, the annotation (2) will have to be done automatically, using a syntax parser with a pipeline producing the annotations in such a way that one can efficiently search and do statistics over the results.

This is something we have, at near-prototype stage.

Once we have that, it is in principle also within reach to develop an automatic detection tool for 'less clear' language, producing alert signals if a text has (many of the) 'less clear'-features.

We will present tools we have in these respects tomorrow.

The technology aside, hoe clear-cut are the supposedly 'non-clear' features – are they in principle well defined, and readily identified in a running text? We will present one case in point, namely what in the English terminology is called *Light Verb Constructions* (*LVC*s) .

Light Verb Constructions (*LVC*s)

One characteristic of *Light Verb Constructions* (*LVC*s) in the sense here addressed is that they unfold, mostly over a sequence '*Subject* **V (P) N'**, a content that could in principle be carried by some verb **V** alone, and where the **N** of the sequence carries the main part of the content, hence the term 'light' for the role of the verb. The **N** thus expresses a situational content, often being 'de-verbal', and a typical role of (the 'light' verb) **V** in the LVC is to connect its *subject* to this situational content as a *role* bearer, and possibly add *aspectual* and *viewpoint* content to the situational content expressed by **N**.

Ex.:

She committed a murder.

The city underwent an attack.

Light Verb Constructions (*LVCs*) - 2

It is conceivable that the LVC might be seen as 'more complex' than a single verb construction, and thus be recommended as not to be used. Without wanting to assess that, we demonstrate what such a phenomenon may represent in terms of 'size' in the language, role in the language, and amenability to analytic and descriptive method. 'Advices' from a linguistic side will have to be grounded in knowledge of all three.

Correspondingly for any other phenomenon subjected to this kind of consideration.

Light Verb Constructions (*LVCs*) - 3

Below are first some examples illustrating the construction type, with the highlighted role indicated, and then a small survey of LVC patterns, as a matter of random choice based on nouns starting with *f*.

In a *Klart språk* initiative, it is conceivable that a 'verb only' alternative would be recommended. The slide subsequent to the next somewhat coarsely illustrates this alternative for the expressions used to illustrate LVCs.

(Most of the 'verb only' examples are a bit contrived, some bordering on ungrammaticality, and some with a loss of expressibility.)

X gjør en feil	'X makes a mistake'	AGENT
X tar et oppgjør med Y	'X takes an issue with Y'	AGENT
X gir inntrykk av Y	'X gives impression of Y'	STIMULUS or REPRESENTATION
X får inntrykk av Y	'X gets impression of Y'	EXPERIENCER
X har en fornemmelse av Y	'X has a feeling of Y'	EXPERIENCER
X får en fornemmelse av Y	'X gets a feeling of Y'	EXPERIENCER
X gir en fornemmelse av Y	'X gives a feeling of Y'	STIMULUS or REPRESENTATION
X foretar et utvalg	'X makes a selection'	AGENT
X begår et mord	'X commits a murder'	AGENT
X undergår et forhør	'X is subjected to an interrogation'	MALEFACTIVE
X gir et tilbud	'X makes an offer'	AGENT
X får et tilbud	'X gets an offer'	RECIPIENT
X mottar en innbydelse	'X receives an invitation'	RECIPIENT
X hengir seg til drikk	'X engulfs in drinking'	AGENT
Det går et rykk igjennom X	'there goes a tremor through X'	PATIENT or LOCUS
X gjennomgår en forandring	'X undergoes a change'	THEME
X gjennomløper en utvikling	'X runs through a development'	THEME
X utfører en operasjon	'X executes an operation'	AGENT
X gjennomfører en undersøkelse	'X conducts an investigation'	AGENT
X tar en jafs av Y	'X takes a bite of Y'	AGENT

X feiler	'X makes a mistake'	AGENT
X gjør opp med Y	'X takes an issue with Y'	AGENT
X virker som Y	'X gives impression of Y'	STIMULUS or REPRESENTATION
X oppfatter Y	'X gets impression of Y'	EXPERIENCER
X fornemmer Y	'X has a feeling of Y'	EXPERIENCER
X fornemmer Y	'X gets a feeling of Y'	EXPERIENCER
X gjør så en fornemmer Y	'X gives a feeling of Y'	STIMULUS or REPRESENTATION
X utvelger	'X makes a selection'	AGENT
X myrder	'X commits a murder'	AGENT
X forhøres	'X is subjected to an interrogation'	MALEFACTIVE
X tilbyr	'X makes an offer'	AGENT
X tilbys	'X gets an offer'	RECIPIENT
X innbydes	'X receives an invitation'	RECIPIENT
X drikker mye	'X engulfs in drinking'	AGENT
Det rykker igjennom X	'there goes a tremor through X'	PATIENT or LOCUS
X forandres	'X undergoes a change'	THEME
X utvikles	'X runs through a development'	THEME
X opererer	'X executes an operation'	AGENT
X undersøker	'X conducts an investigation'	AGENT
X jafser av Y	'X takes a bite of Y'	AGENT

LVCs with nouns starting with *f*

39 verbs, 258 nouns, partaking in 250 AG-profiled LVCs and 110 THEME/PAT-profiled LVCs

Head verb	English trans.	*Eigen*-roles	LVCs	Sample nouns
foreta	conduct, do	Ag	72	fengsling, forbyttelse, fordrivelse, fordeling, forenkling, forfremmelse, forskuttering, forsøk, fortolling, frakobling, frigjøring, forsøpling, fortetning
være/bli gjenstand for	be subject to	Pat, Th, Ben, Mal	67	forulempelse, forurettelse, forutbestemmelse, fortielse, forslumming, forsøk, forsøpling, fortetning, forundring
drive	conduct	Ag	52	forskning, forvaltning, filosofering, fotografering, fragmentering, falsifikasjon, fordummelse, forherligelse, forskjønnelse
undergå	undergo	Pat, Th	44	fortielse, forslumming, forsøk, forsøpling, fortetning, forvandling, forvitring
begå	commit	Ag	28	feilvurdering, forbrytelse, fortielse, fornærmelse, forstyrrelse, fusk
gi	give	'Initiator'	17	forlatelse, fritak, forklaring, forelesning, forestilling, forordning, fortolkning, fremføring, fornemmelse

LVCs with nouns starting with *f*
39 verbs, 258 nouns, partaking in 250 AG-profiled LVCs and 110 THEME/PAT-profiled LVCs

Head verb	English trans.	*Eigen*-roles	LVCs	Sample nouns
gjøre	do	Ag	16	fangst, fortjeneste, fritak, feilvurdering, foranstaltning, forbedring, forfalskning, forsøk, forberedelse
ha	have		11	forbruk, forståelse, formening, fornemmelse, følelse, forankring, forløp
utløse	release, cause	Ag	9	forargelse, forbarmelse, forbauselse, forbitrelse, forbløffelse, forsinkelse, forferdelse, forskrekkelse
få	get	Rec, Ben, Mal	7	flyt, forløsning, forståelse, fornemmelse, følelse, forankring
hengi seg til			4	forlystelse, fornøyelse

The study of LVCs

Many phenomena have been chracterized in terms of the notion 'light verb'; Butt 2010 may be seen for a summary of many of them. The usage of the term here employed goes back at least to Jespersen 1964, more recently Grimshaw and Mester 1988, and is a topic of much current attention, see, e.g., Piunno & Pompei 2015, Nagy et al. 2013. (Other uses of the term, found not least in the Scandinavian oriented literature, pertain to auxiliary-like items (Lødrup 2002), and to first verbs in so-called pseudo-coordination (cf. Brøcker 2013).) It is well recognized that LVCs as here understood constitute a major category in Persian and in Indic languages, and likewise in West African languages; the present paper is concerned with Norwegian only, and presents what may seen as a 'reconnaissance' for a larger project on LVCs in the language, offering an appraisal of the general scope of LVCs in the language, a diagnosis of the 'gist' of the construction, and a framework for formal analysis attuned to this diagnosis, comprising both an analytic architecture suitable for formal grammars and a notation format for constructing anotated corpora of LVCs.

Interim assessments

(i)   *No verb per se is defined as a 'light verb'.*

Any verb used in an LVC has a wide field of use other than that of LVCs, and to address this fact, the lexical entry versions accommodating the various uses of the verb should ideally be as close as possible. By maintaining the general argument structure of verbs in the LVC-related lexical entries, the analysis should address this concern.

(ii)   *No noun is per se defined as a 'light verb dependent' noun.*

It is a prerequisite for a noun to take part in an LVC that it has a situational content, but apart from the reference to this property through a situational specification (call it *SIT*), the analysis should not presuppose any special status of the noun.

(iii)   *No pattern of selection is defined as a specific 'light verb selection' pattern.*

Apart from the circumstance that the LVC-related lexical entries state a certian SIT identity between the verb semantics and the noun semantics, the mechanisms otherwise deployed in the selectional rule of the verb relative to the noun or the PP should follow the general patterns of such statements.

Main generalizations concerning LVCs

As for their build-up, their main feature of LVCs is presumably that verb and noun cooperate in highlighting a semantic role, and thus that verb and noun in some sense *share* that role.

There is then the question why there are LVCs at all. One can conceive that a situational content in seeking to get a linguistic expression has more than one channel of gaining such expression. The category of *verb* is one channel, bound to the patterns of valency offered by the grammar of the language. Another is through the category of *noun*, whereby reference to the situational content for uses of quantification, definiteness specification and other can be provided, and, through embedding in an LVC, the LVC can come to serve like one more possible pattern of valency relative to the situational content.

Excerpt of a possible situation-type hierarchy
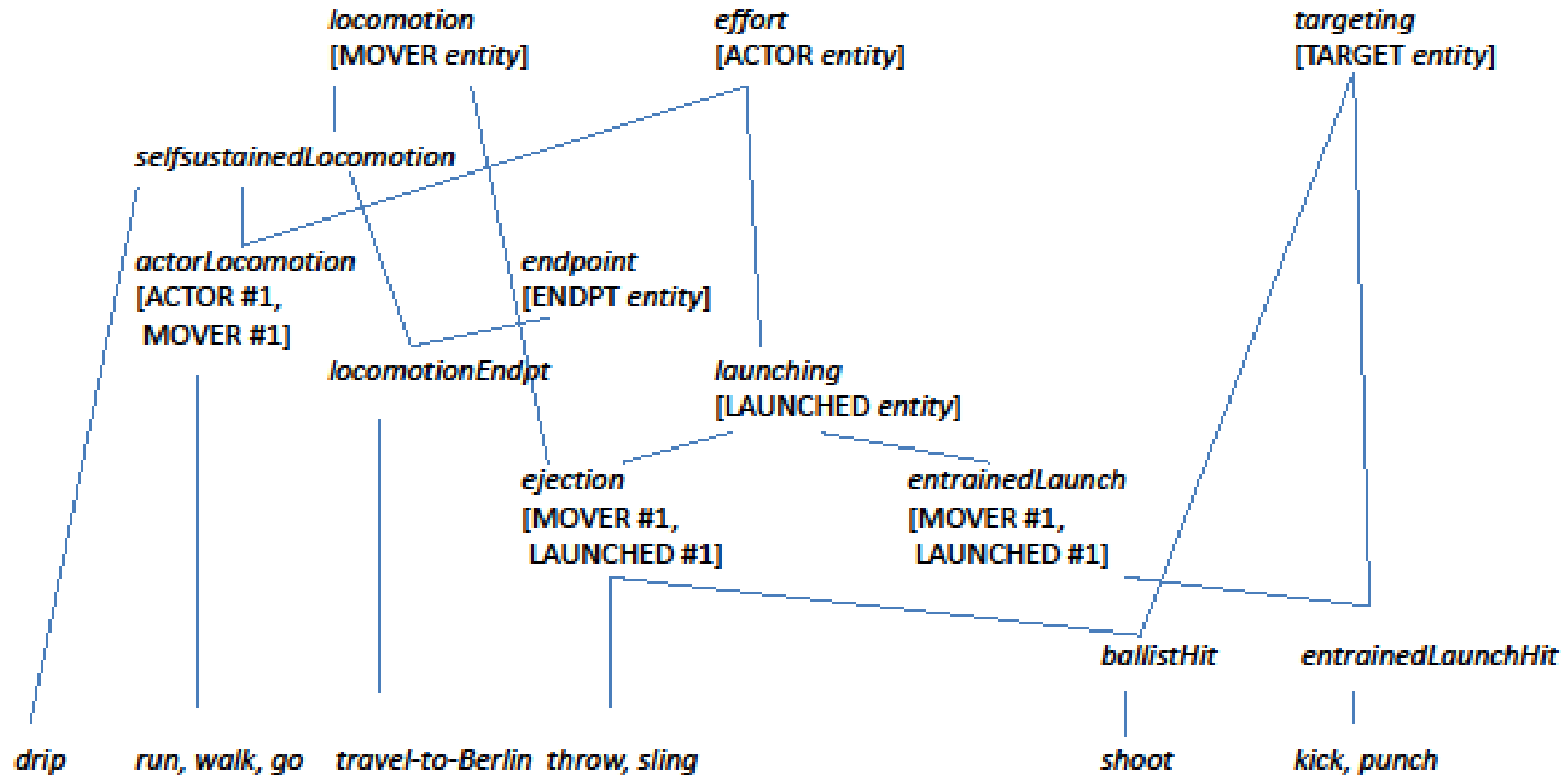
LVC composition

We now turn to the composition of LVCs, and unification relating to semantic types of the verb and the noun. As an example, we use

Banken foretok en nedskrivning av kronen
'The bank performed a devaluation of the 'krone''

Assuming that nouns lack argument structure, unification will have to relate to SIT content:
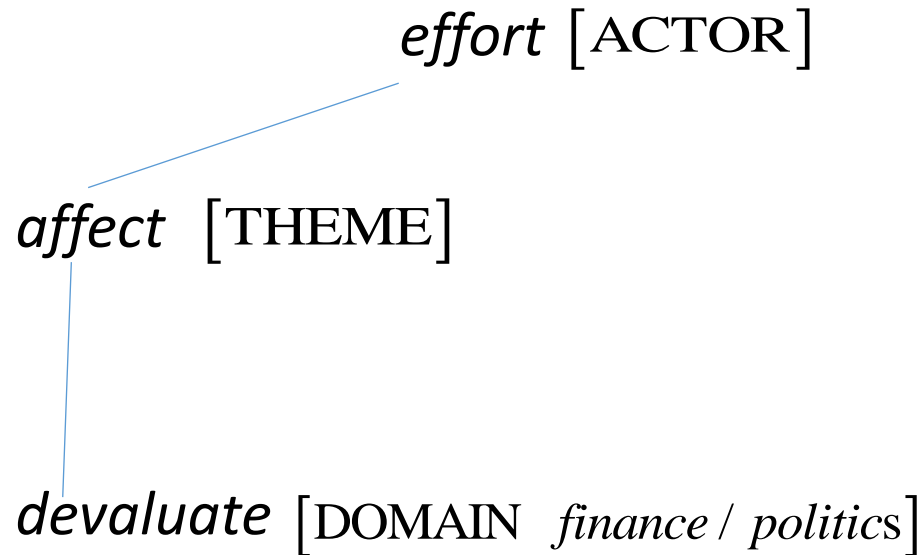
foreta:

$$\begin{bmatrix} \text{HEAD verb} \\ \text{STYLE formal} \\ \\ \text{SIT } \textit{effort} \begin{bmatrix} \text{ACTOR} \end{bmatrix} \end{bmatrix}$$

nedskrivning:

$$\begin{bmatrix} \text{HEAD noun} \\ \text{STYLE formal} \\ \\ \text{SIT } \textit{devaluate} \begin{bmatrix} \text{ACTOR} \\ \text{THEME} \\ \text{DOMAIN } \textit{finance / politics} \end{bmatrix} \end{bmatrix}$$

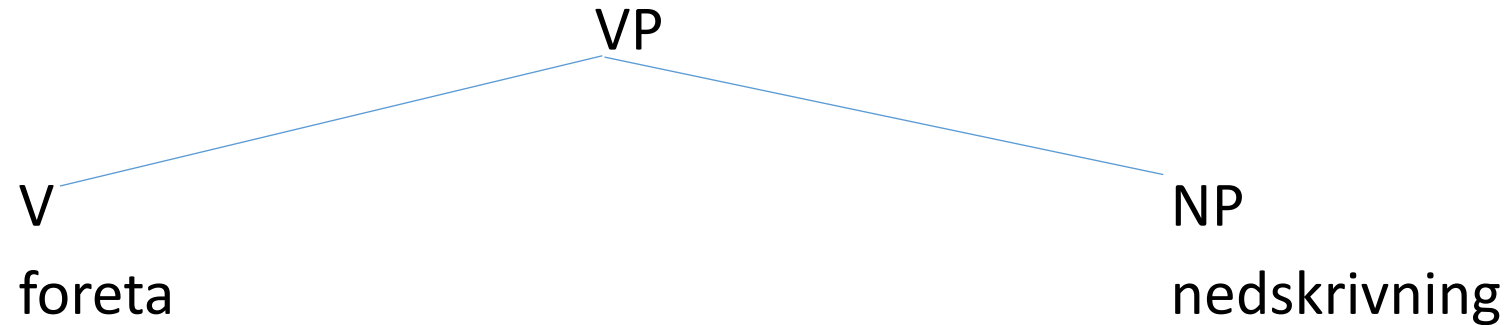# Compatible SITs

We want to represent *foreta* and *nedskrivning* as having matching SITs. Adding other inheritance lines under *effort* in the hierarchy shown earlier, the situation types *effort* and *devaluate* can be construed as **type compatible**:

$$effort \; [\textsc{actor}]$$

$$affect \; [\textsc{theme}]$$

$$devaluate \; [\textsc{domain} \; \textit{finance} \,/\, \textit{politic}s]$$

Composition of LVC

For the VP of the sentence mentioned, the phrase structure will be essentially:

```
                    VP
                   /  \
                  /    \
         V                      NP
       foreta                nedskrivning
```

For this combination of signs, the complement in the valency frame of *foreta* – introduced by the attribute 'OBJ' - will be defined as SIT compatible with the verb itself, reflecting the status of *foreta* as a 'light verb'. This is expressed through the identity-marking [1]-boxes in the following AVM-snippet:

$$
\begin{bmatrix}
\text{GF}\begin{bmatrix} \text{OBJ}\begin{bmatrix} \text{SIT} & \boxed{1} \end{bmatrix} \end{bmatrix} \\
\text{SIT} \ \boxed{1} \ \textit{effort}
\end{bmatrix}
$$

Since the situation type *devaluate* is a subtype of *effort,* this compatibility condition is met, and the phrase *foreta nedskrivning* is thus accepted by the grammar.

Selection

As witnessed by the circumstance that *foreta* can appear in no less than 72 LVCs where the noun starts with f, this is a 'light verb' with a highly general distribution; in this, it is rather untypical, since most LVC combinations are rather idiosyncratic. One example is *lide nederlag* 'suffer defeat': here *lide* is among the very few 'light' verbs that can combine with *nederlag*, and vice versa. To build such a circumstance into the combination formalism, not only POS and SIT of the complement must be specified in the verb's valence frame, but also some sort of sign-specific identification of the object noun. Schematically, this will look like the following AVM snippet from the sign for *lide* (where the the attribute 'KEY' serves for sign identification; note that SIT identity now concerns a PATIENCE role):

$$
\begin{bmatrix}
\text{GF} \begin{bmatrix} \text{OBJ} \begin{bmatrix} \text{HEAD}\begin{bmatrix} \text{KEY } \textit{nederlag} \end{bmatrix} \\ \text{SIT } \boxed{1} \end{bmatrix} \end{bmatrix} \\
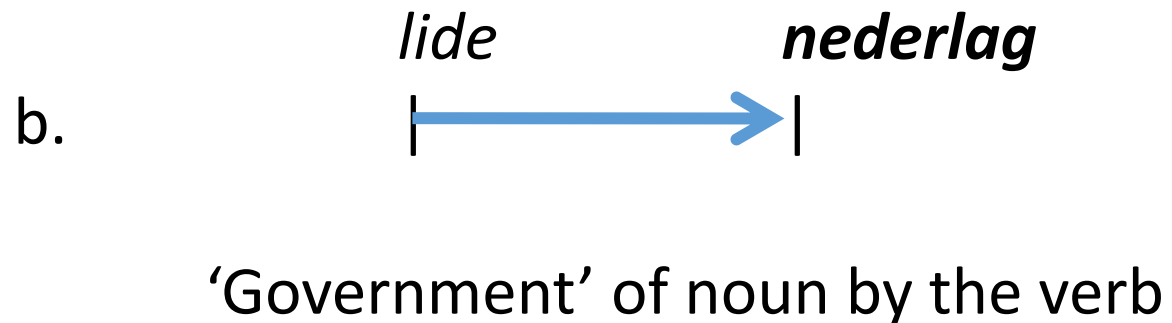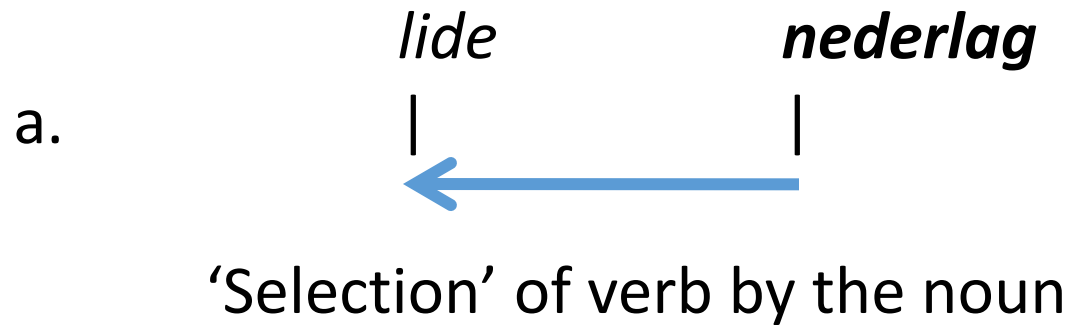\text{SIT } \boxed{1}
\end{bmatrix}
$$

We exemplify how a more complete view of such an AVM will look, including ACTs:

Lide 'suffer' selecting *nederlag*

$$
\begin{bmatrix}
\text{ORTH} \langle \textit{"lide"} \rangle \\
\text{HEAD} [\textit{verb}] \\
\text{GF} \begin{bmatrix}
\text{SUBJ} \begin{bmatrix} \text{INDX} \begin{bmatrix} \text{HAEC} \boxed{3} \end{bmatrix} \end{bmatrix} \\
\text{OBJ} \begin{bmatrix}
\text{HEAD } \textit{noun} \ [\text{KEY } \textit{nederlag}] \\
\text{ACTNTS} \begin{bmatrix} \text{ACT0} \begin{bmatrix} \text{HAEC} \boxed{6} \end{bmatrix} \end{bmatrix} \\
\text{SIT} \boxed{1}
\end{bmatrix}
\end{bmatrix} \\
\text{ACTNTS} \begin{bmatrix}
\text{ACT1} \begin{bmatrix} \text{HAEC} \boxed{3} \end{bmatrix} \\
\text{ACT2} \begin{bmatrix} \text{HAEC} \boxed{6} \end{bmatrix}
\end{bmatrix} \\
\text{SIT} \boxed{1}
\end{bmatrix}
$$

Selection

The above AVM portrays the selection in an LVC as something excercised by a verb over the complement. However, given that the noun equally much selects the verb, one might conceive of the selection construed the opposite way, as schematically indicated in (a), whereas the standard construal is indicated in (b); this option we leave open:

lide   **nederlag**

a.

'Selection' of verb by the noun

lide   **nederlag**

b.

'Government' of noun by the verb

Towards corpus annotation for LVCs

To create an analytically telling corpus of LVCs, one needs an exhaustive classification of the types that may appear, which will mean, an exhaustive classification of the types of situational nouns that can be encountered, and of the types of possible 'light verbs'. The classifications should correspond to the categories used in the grammar, thus in the AVMs, and the annotation tags and patterns should be recursively convertible into the grammar formalism.

We illustrate the issue with a tentative classification and tag system for nouns, and its deployment in an annotation for LVCs.

Parameters for classification of nouns

Parameter	Parameter abbreviation	Values	Value abbreviations
Ontological status	Ontstat	Situation vs. Thing	s, t
Resultativity	Res	Result of event vs. not	1, 0
Agentivity	Ag	Agentive vs. Non-agentive	1, 0
Aspect	Asp	Aspectual types	Type name(s) (connected with '&' when many)
Institutionalization	Instit	Institutionalized vs. not	1, 0
Domain	Dom	Physical vs. Cognitive vs. Emotional vs. Apriori vs. Social vs. *FinanJurAdminManag* (=fjam)	phys, cog, emot, aprio, soc, fjam (connected with '&' when many)
Valency preservation	Val	Valency preserving vs. not	1 - 0
Theta-role, for things	Th	The role that the entity has relative to the sit-type expressed by the root	Role name(s) (connected with '&' when many)

Their application to classification of nouns can go as indicated below:

		Ontstat	Res	Ag	Asp	Instit	Dom	Val	Th
bønn_s	'prayer'	s	0	1	dur	1	cog	0	
bønn_t	'prayer'	t	0	1	--	1	cog	0	inh
begjær	'desire'	s	0	0	dur	0	emot	0	
begrep	'concept'	t	0	0	-	1	cog	0	inh
behag	'pleasure'	s	0	0	dur	0	emot	0	
behov	'need'	s	0	0	dur	0	All	1	
besøk	'visit'	s	0	1	dur	&&	soc&fjam	0	
bifall	'approval'	s	0	1	dur	&&	soc&fjam	0	
bistand	'support'	s	0	1	all	&	fjam	0	
bitt	'bite'	t	1	1	-	0	phys	0	inh
brak	'crash'	s	&&	0	inst	0	phys	0	
brann	'fire'	s	&&	0	dur	0	phys	0	

For annotation of noun occurrences in a corpus, one can in turn pull such value sequences together in short-hand expressions, as indicated below in an annotation snippet for a construction including the light verb expression *lide nederlag* 'suffer defeat', the shorthand reflecting the above stated values for *nederlag*, marked with POS value 'BN' for 'bare nominalization', and with 'LVC1' marking the LVC.

lide	*nederlag*
suffer	defeat
LVC1	LVC1
V	**BN**=s10inst0All0

Annotation tags

Formulaic expressions like `BN=s10inst0All0` are easy enough to master for an annotator, they are easy to expand into readable prose if one wants to, and they can be converted into grammar code, like AVMs like those exemplified above, for the purpose of grammar induction from corpus annotation. For instance, the occurrences of 'LVC1' (indicating 'catenae' – cf. Osenova and Simov 2015) will mean that the words with this annotation will be identity-reflected in an AVM like the last one shown above, and the marking of *lide* as head means that it will represent the selecting head in such an AVM. (Notice that by itself, this 'LVC1' marking is neutral regarding what is construed as selecting head - cf. slide above.)

Annotation tags

The string `s10inst0All0` can be processed by a finite-state-like procedure building up an AVM for the noun where 's' determines the type as situational, '1' yields a specification as an outcome, '0' prevents the appearance of an ACTOR attribute, 'inst' gives the aspectual feature 'Protracted –', etc. (One could alternatively use order-independent symbols which each would correspond to a partial AVM, and whose combination would be interpreted as unification of these AVMs.)

We thus succeed in tying corpus annotation and formal grammar together.

Annotation in corpus

The purpose of having an annotated corpus is partly to have an accessible 'example bank' for the phenomenon in question, partly - and more significantly - to have a *proof of existence* of the phenomenon. The latter requires a corpus with clear and tractable metadata. For Norwegian, the possibility for obtaining such corpora at a large scale is about to become available through the (p.t.) 15 billion words corpus at *NB-digital,* the National Library's assembly of digital texts. As an indication of this facility, we enter below the number of concordances for a small set of LVC strings found in this corpus.

Annotation in corpus

It may be noted that these strings of words were defined by the author, and submitted to the database for concordance search. Where the search responds with a reasonable number of concordances, and inspection shows that the match has indeed the intended phrasal structure (and the metadata are ok), this is a true proof of existence. From among the data found, one can in turn select some and enter into another corpus to be subjected to annotation, with metadata tracing the source in NB-digital.

This procedure is different from one where in a large corpus one tries to automatically detect LVCs - see, for instance, Nagy et al. 2013, Grefenstette andTeufel 1995. Neither procedure is by itself more 'empirical' than the other, as long as metadata-confirmed data in the end is what supports the analyses. And given the lack of formal distinguishing features of LVCs, designing such a procedure will be a challenge.

Corpus of LVCs?

The following are the number of concordances for a small set of LVC strings found in a 15 billion words corpus in *NB-digital*, the National Library's assembly of digital texts:

'gi forklaring' – 4140,

'gi forlatelse' - 50,

'foreta fordeling' - 240,

'ilegge forelegg' - 45,

'inngå forlik' - 1500,

'drive forskning' – 330,

'vise forakt' - 660,

'få forståelse' - 4800,

'holde foredrag' – 5000+,

'være gjenstand for' – 5000+,

'inngå forlovelse' - 80,

'forhold opptrer' – 100,

Conclusion - 1

The following are questions which our model will allow us to formulate:

(i) For a given SIT profile, are there principles determining which valency frame (of the language in question) may be used to support its realization as a verbal sign?

(ii) For a given SIT profile, are there principles determining which LVC pattern (of the language in question) may be used to support its realization as a nominal sign, and with which verbs?

Thus, just as (i) invites to a strategy where valency frame assignment can be made automatic rather than proceed verb by verb, so (ii) invites to a strategy where from the meaning of a situational noun one can predict with which verbs it can constitute an LVC.

We have situated the analysis relative to a partial overview of LVCs in Norwegian, with the aim of conducting a large scale investigation both in grammar implementation and in corpus encoding. At the same time, the analytic design may well be applicable across all languages containing the phenomenon, and serve as a frame within which comparison across languages can in turn be made.

Conclusion - 2

Comparison with German is an obvious such task.

As for *Klart Språk*, the paper exemplifies a phenomenon that might be 'alert-marked' in a guideline, but which is obviously not one to try to simply avoid. In this, LVCs potentially aligns with phenomena such as

- Passive

- Anaphora

- Compounding

- Subordination
    - Embedding within embedding
    - Concessive constructions

Conclusion - 3

If we want to build an automatic 'alerter' for various phenomena, that can in principle be built on to a Grammar Checker (more on that tomorrow).

As we have seen, there is no overt formal marking of LVCs, such that for LVCs to be present in a parser, we will have to list them all, and perhaps specially mark those that are unduly heavy.

For compounding it is likewise. There are no general traits by which they can be recognized (except through a word-internal parser saying that there are independent lexical items constituting the word), and even if, most compounds are totally common in the language, so that a special list would have to be made of those requiring specific alertness. In this case such a list may well be domain specific, and so a matter of institutional *oppdragsforskning* rather than something residing in a general linguistic resource.

Either way, a lot of work - - so it better be interesting.

References

Beermann, D. (2014). Data management and analysis for less documented languages. In Jones, M., and Connolly, C. (eds) *Language Documentation and New Technology*. Cambridge University Press.

Butt, Miriam (2010). "The Light Verb Jungle: Still H acking Away". In: Complex Predicates in Cross-Linguistic Perspective. Ed.by M. Amberber, M. Harvey, and B. Baker. Cambridge: Cambridge University P ress,pp.48–78.

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

Davis, Anthony. 2000. *The Hierarchical Lexicon*. CSLI Publications, Stanford.

Grefenstette, G. and S. Teufel 1995. "Corpus-based method for automatic identification of support verbs for nominalization". In Proceedings of the 7th Meeting of the European Chapter of the Association for Computational Linguistics (EACL'95).

Grimshaw, Jane, and Armin Mester. (1988). Light Verbs and Ɵ-Marking. *Linguistic Inquiry* 19(2):205–232.

Hellan, L. 2010. From Descriptive Annotation to Grammar Specification. *Proceedings of the Fourth Linguistic Annotation Workshop (LAW4)*, 2010, ACL, pp 172--176, http://www.aclweb.org/anthology/W10-1826.

References contd

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

Hellan, L. and M. E. Kropp Dakubu, 2010: *Identifying Verb Constructions Cross-Linguistically. Studies in the Languages of the Volta Basin* 6.3. Legon: Linguistics Dept., University of Ghana. http://www.typecraft.org/w/images/d/db/1_Introlabels_SLAVOB-final.pdf.

Jespersen, Otto. 1965. *A Modern EnglishGrammar on Historical Principles, Part VI, Morphology.* London: George Allen and Unwin Ltd.

Levin, B. & Rappaport Hovav, M. (2005). *Argument realization,* Cambridge: Cambridge University Press.

Lødrup, Helge (2002)."The syntactic structures of Norwegian pseudocoordinations". In: StudiaLinguistica 56.2,pp.121–143.

Melchuk, I. 2004. Actants in semantics and syntax I: actants in semantics. *Linguistics* 42-1: 1-66.

István Nagy T., Veronika Vincze, and Richárd Farkas. (2013) Full-coverage Identification of English Light Verb Constructions. International Joint Conference on Natural Language Processing, pages 329–337, Nagoya, Japan, 14-18 October 2013.

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

Pompei, A. and V. Piunno (2015). Light Verb Constructions. An interlinguistic analysis to explain systemic irregularities. Paper presented at SLE 2015, Leiden.

Tanja Samardzíc and Paola Merlo. 2010. Cross-lingual variation of light verb constructions: Using parallel corpora and automatic alignment for linguistic research. In Proceedings of the 2010 Workshop on NLP and Linguistics: Finding the Common Ground, pages 52–60, Uppsala, Sweden, July. ACL.