

**Light Verb Constructions as *valency modeling*. A study of Norwegian.**

**Lars Hellan**

NTNU

SLE, Naples August 31, 2016

## LVC - characterization

One characteristic of *Light Verb Constructions* (LVCs) 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**.

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*.

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

# 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.	<i>Eigen</i> -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.	<i>Eigen</i> -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

Many phenomena have been characterized 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 be 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 annotated 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 certain *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.

Without making a teleological issue of the latter, we will outline a formal analytic architecture allowing for the modeling of hierarchies of situation types and how they can be integrated in grammatical description so that they can be expressed through verbs as well as nouns. This framework of grammatical description then also will allow us to model the ‘cooperation’/’sharing’ between the verb and the noun in an LVC.



# Situation types

Situational contents may be conceived as organized in type hierarchies, where types can introduce attributes for further specification, and inherit from multiple higher types. The two following slides illustrate the idea, the second slide combining Aktionsart types with situational types. The design follows these principles (laid out in Copestake 2002):

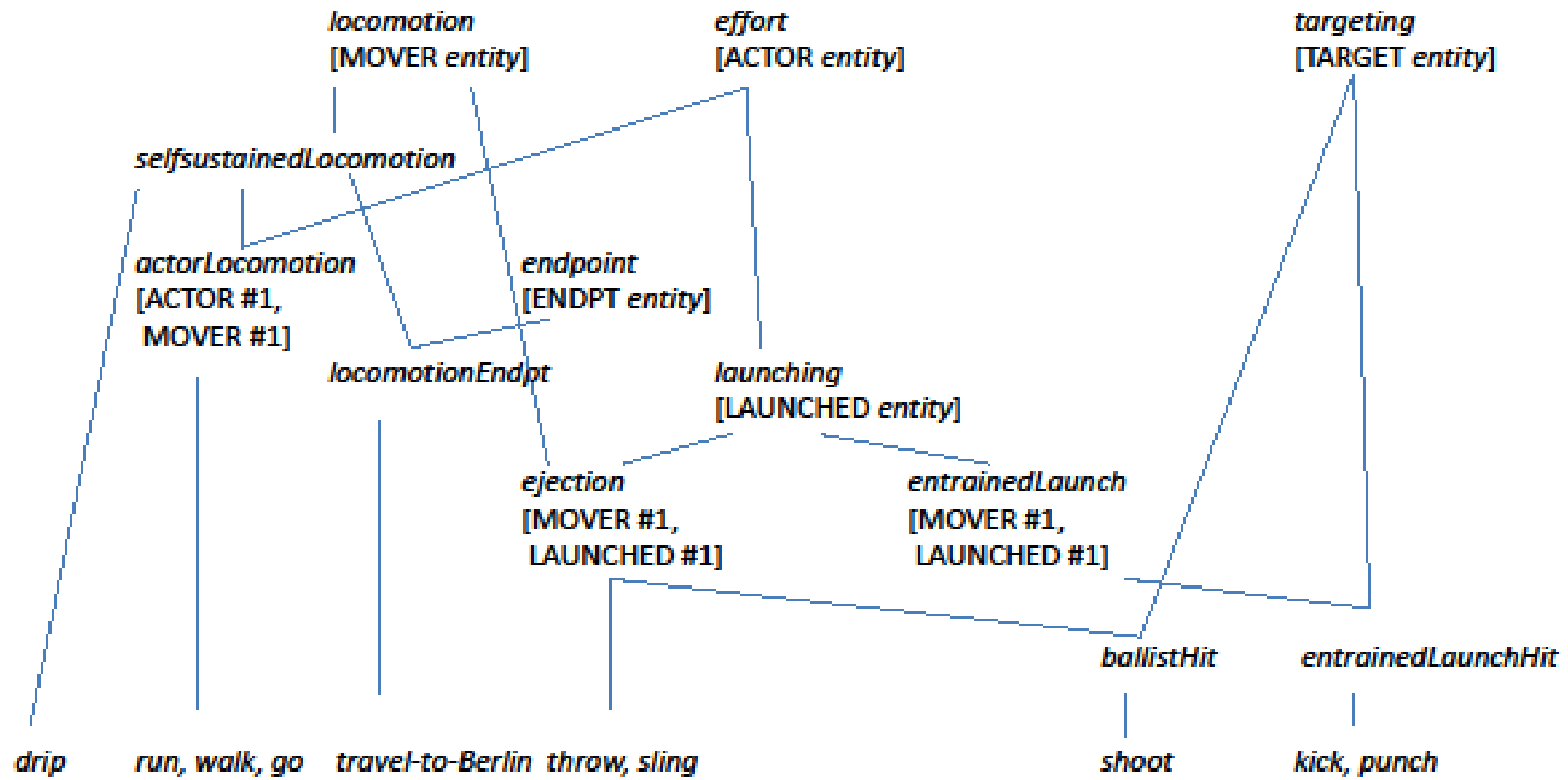
[A] A given type introduces the same attribute(s) no matter in which environment it is used.

[B] A given attribute is declared by one type only (but occurs with all of its subtypes).

This style of attribute-enriched type hierarchy is used, e.g., in Davis (2000), although with different types and attributes.

(The below illustrations in some cases leave specifications incomplete, like a node for 'COMPLETED –', and thus whether 'run' is completed or not – lower type to this node can of course be either, depending on inheritance from the relevant aspectual types.)

# Excerpt of a possible situation-type hierarchy



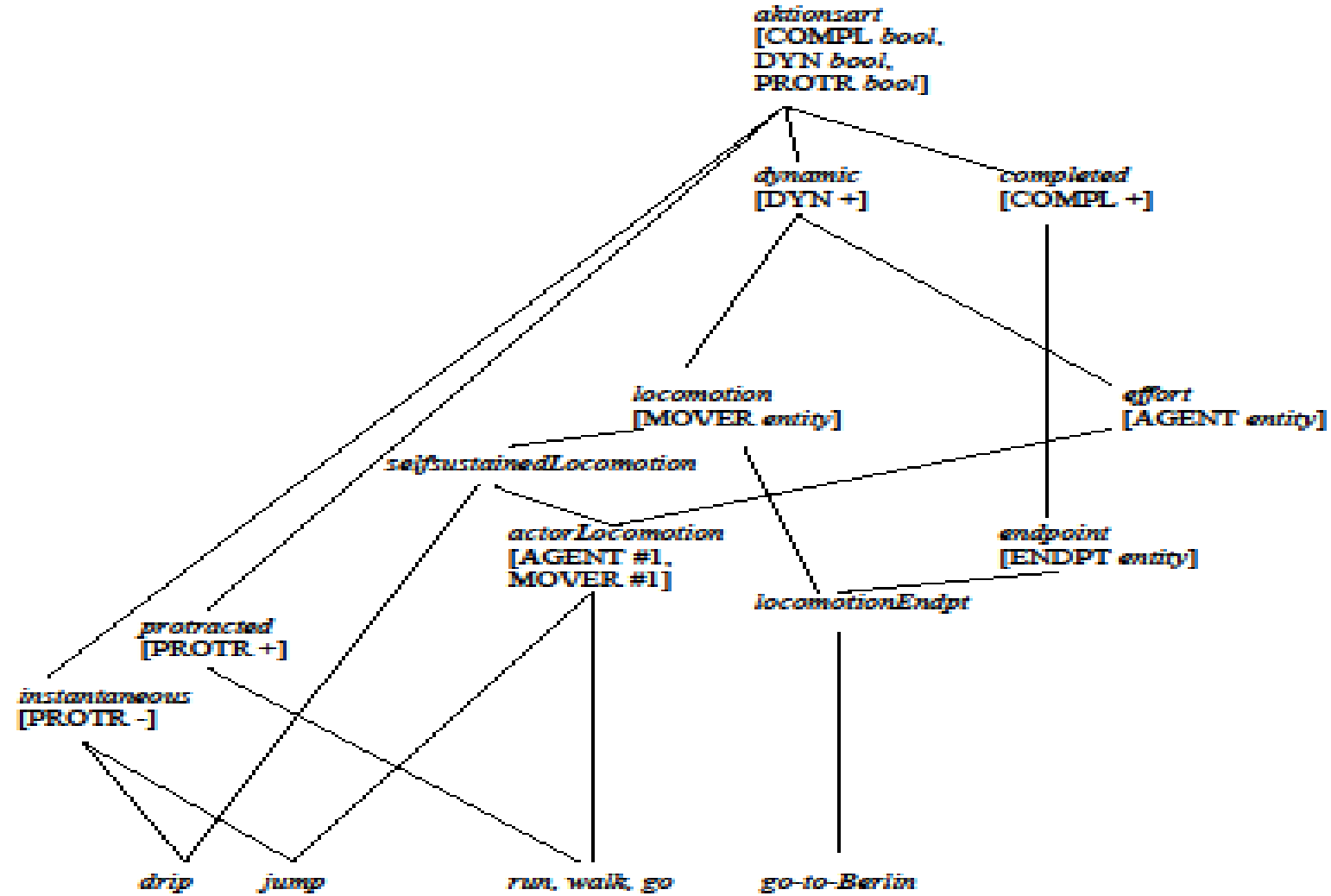


Figure 7 Partial combination of type hierarchies for Aktionsart and situation types (with incomplete drawing of possible super- and sub-types)

# Grammar integration

We next outline how situational content can be built into signs for both verbs and nouns, and how situational content relates to verbal argument structure. The first slide illustrates how a 'root' sign can be modeled as expanding to either a verb or a noun. We argue that a verb's semantic argument structure is distinct from its situational content, the subsequent slides reflecting on what kind of information resides in the respective components.

After that, we turn to the composition of LVCs, and the role of unification relating to semantic types of the verb and the noun.

# Situational content in lexical sign inheritance

**Root:**

*SIT* + rudimentary morpho-phonological specification

----- [*Type inheritance*] -----

**Verb entry**

**+ *ACTNTS* + *GF***

**Bare nominalization entry**

# Semantic argument structure

The *semantic arguments* expressed by the sentence can be introduced by an attribute such as **ACTNTS**, being distinguished as ‘actant 1’ (**ACT1**), ‘actant 2’ (**ACT2**), etc. The **GF** and **ACTNTS** values can be interlinked through the individuals serving as *bearers* of the *actant* functions, identified by a pointer entered as value of the **ACT** attribute, which can at the same time be seen as the *referents* – introduced by the attribute **INDX** – of the grammatical functions. The **GF** and **ACTNTS** values thereby together give a representation of a *sign*. The AVM below can for instance represent a sentence like *John kicks Peter*:

$$\left[ \begin{array}{l} \text{GF} \left[ \begin{array}{l} \text{SUBJ} \left[ \text{INDX } \boxed{1} \right] \\ \text{OBJ} \left[ \text{INDX } \boxed{2} \right] \end{array} \right] \\ \text{ACTNTS} \left[ \begin{array}{l} \text{ACT1 } \boxed{1} \\ \text{ACT2 } \boxed{2} \end{array} \right] \end{array} \right]$$

# Root

[ ORTH <"løp">  
SIT *run* [ ACTOR 3  
MOVER 3  
PROTR +  
DYN + ] ]

# Verb

[ ORTH <"løpe">  
HEAD [*verb*]  
GF [ SUBJ [ INDX. 1 ] ]  
ACTNTS [ PRED *løpe \_rel*  
ACT1 1 [ HAEC 3  
ROLE *locomotor* ] ]  
SIT *run* [ ACTOR 3  
MOVER 3  
PROTR +  
DYN + ] ]

# Noun

[ ORTH <"løp">  
HEAD [*noun* [ GENDER *neut* ] ]  
SIT *run* [ ACTOR 3  
MOVER 3  
PROTR +  
DYN + ] ]

## A 'shadow' of syntactic structure ...

Our proposal is that only **verbs**, and *nouns regularly derived from verbs*, have a semantic argument structure – it sits a bit like a 'shadow' of the syntactic structure, although not identical to it. Ways in which it can deviate from syntactic structure are illustrated in the following AVMs:

*'Expletive subject':* 
$$\left[ \begin{array}{l} \text{GF} \left[ \begin{array}{l} \text{SUBJ} \left[ \text{HEAD pron} \right] \\ \text{OBJ} \left[ \text{INDX } \boxed{1} \right] \end{array} \right] \\ \text{ACTNTS} \left[ \text{ACT1 } \boxed{1} \right] \end{array} \right]$$

*'Implicit object':* 
$$\left[ \begin{array}{l} \text{GF} \left[ \text{SUBJ} \left[ \text{INDX } \boxed{1} \right] \right] \\ \text{ACTNTS} \left[ \begin{array}{l} \text{ACT1 } \boxed{1} \\ \text{ACT2 index} \end{array} \right] \end{array} \right]$$

*'Blocked subject role':* 
$$\left[ \begin{array}{l} \text{GF} \left[ \text{SUBJ} \left[ \text{INDX } \boxed{1} \right] \right] \\ \text{ACTNTS} \left[ \begin{array}{l} \text{ACT1 blockindex} \\ \text{ACT2 } \boxed{1} \end{array} \right] \end{array} \right]$$



## The ACT attributes

The attributes **ACT1**, **ACT2** etc. as used here are partly *role labels*, and partly *enumeration markers*: As enumeration markers, they list the participants present in the situation expressed (including implicit ones), starting with ACT1, using ACT2 only if there is an ACT1, and using ACT3 only if there is an ACT2. (This is analogous to the conventional listing of arguments of an operator in logical notation, where in expressions like 'P(x,y)' one introduces a comma only if there is more than one argument.) As role markers, when there is more than one argument, they express something close to 'macro' or 'proto' roles, so that when there is an ACT1 and an ACT2, ACT1 is the role associated with emanation of force, and ACT2 is the 'impacted' part relative to the force; an ACT3 would then express a slightly less directly involved participant than the ACT2, such as the recipient or benefactive in a ditransitive sentence.

## The ACT attributes - 2

When there is only one actant, it will be marked as ACT1, regardless of its role. (Again, this is analogous to conventional logical notation.) Still, the *ACTn* attributes are not mere replica of the GFs of the sentence represented: apart from the circumstance that also implicit participants receive an *ACTn*, the ordering among the *ACTn* s does not necessarily reflect the actual GFs carried by the constituents expressing the participants in question. For instance, although in a sentence like *the apple was eaten*, *the apple* is a subject, it will correspond to the ACT2 participant.

# Roles in semantic argument structure

Distinguishing between no more than three (or four) participant types, the *ACT<sub>n</sub>* attributes by no means purport to fully differentiate between all types of roles that can be recognized. To increase role expressiveness, while keeping the overall frame of the ACTNTS argument structure, one can introduce ROLE as an attribute inside of *ACT<sub>n</sub>*, with explicit *role* notions as value; these still will constitute a far more restricted space than the SIT roles:

Representing roles in argument structure:

$$\left[ \text{ACTNTS} \left[ \begin{array}{l} \text{ACT1 index [ROLE ag]} \\ \text{ACT2 index [ROLE theme]} \end{array} \right] \right]$$

# 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:*

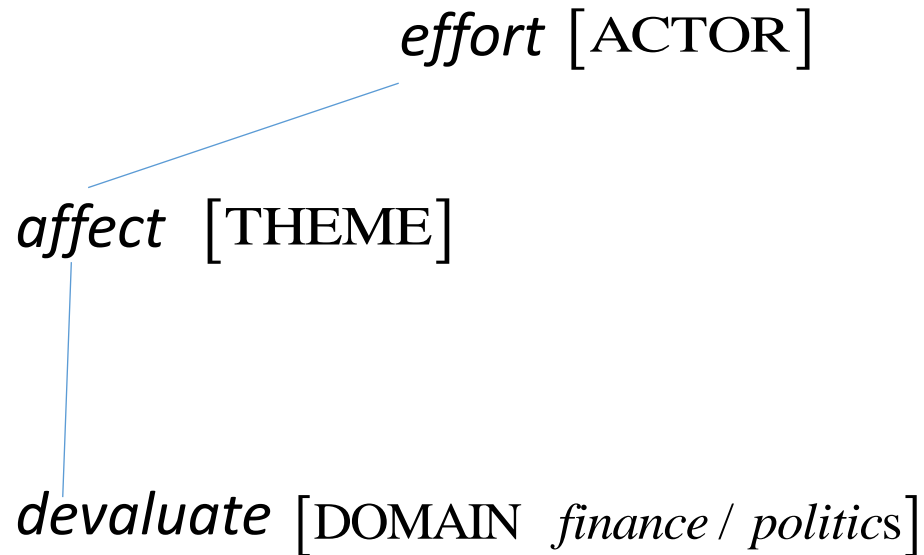
[ HEAD verb  
STYLE formal  
SIT *effort* [ACTOR] ]

*nedskrivning:*

[ HEAD noun  
STYLE formal  
SIT *devalue* [ ACTOR  
THEME  
DOMAIN *finance / politics* ] ]

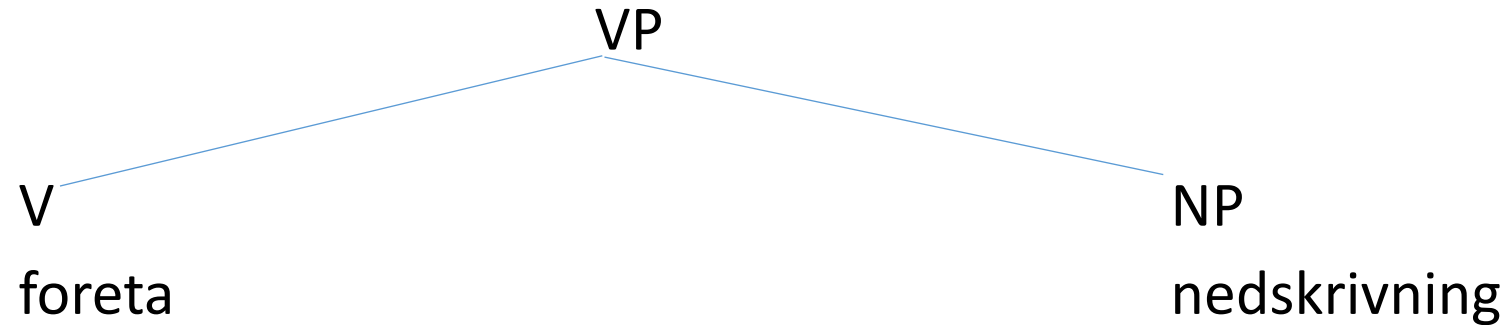
# Compatible SITs

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 *devalueate* can be construed as ***type compatible***:



# Composition of LVC

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



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:

$$\left[ \begin{array}{l} \text{GF} \left[ \text{OBJ} \left[ \text{SIT} \boxed{1} \right] \right] \\ \text{SIT} \boxed{1} \textit{effort} \end{array} \right]$$

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 valency 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 attribute ‘KEY’ serves for sign identification; note that SIT identity now concerns a PATIENCE role):

$$\left[ \begin{array}{l} \text{GF} \left[ \text{OBJ} \left[ \begin{array}{l} \text{HEAD} [\text{KEY } \textit{nederlag}] \\ \text{SIT } \boxed{1} \end{array} \right] \right] \\ \text{SIT } \boxed{1} \end{array} \right]$$

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

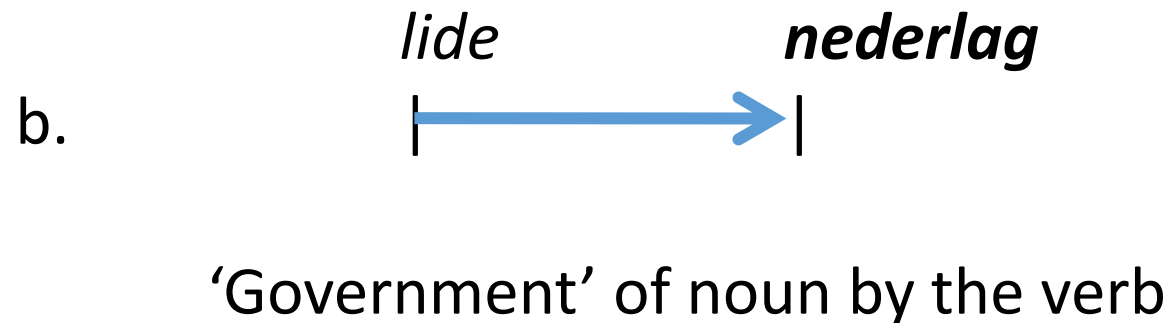
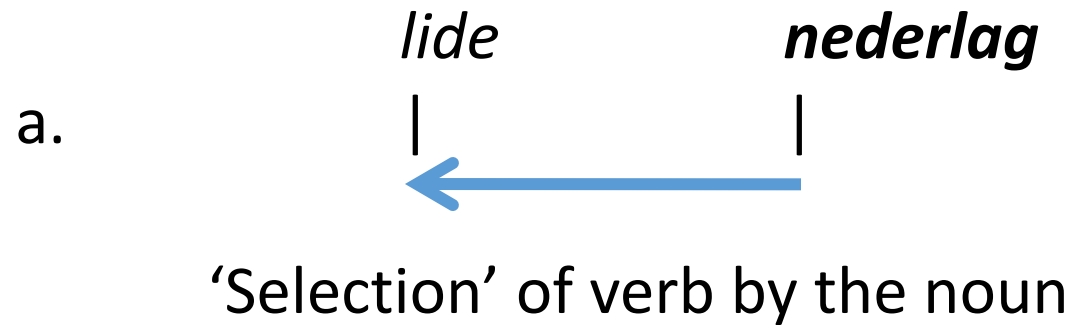
# *Lide* 'suffer' selecting *nederlag*

[ ORTH <"lide">  
HEAD [verb]  
GF [ SUBJ [ INDX [ HAEC 3 ] ]  
OBJ [ HEAD noun [ KEY nederlag ]  
ACTNTS [ ACT0 [ HAEC 6 ] ]  
SIT 1 ] ] ]  
ACTNTS [ ACT1 [ HAEC 3 ] ]  
[ ACT2 [ HAEC 6 ] ]  
SIT 1 ] ] ]



# Selection

The above AVM portrays the selection in an LVC as something exercised 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:



# 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, for instance, for the purpose of grammar induction from corpus annotation (cf. Hellan 2010, Beermann 2013, Beermann and Hellan 2014).

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. <i>FinanJurAdminManag</i> (=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.

<i>lide</i>	<i>nederlag</i>
suffer	defeat
LVC1	LVC1
V	BN=s10inst0A110

## Annotation tags

Formulaic expressions like `BN=s10inst0A110` 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.) The string `s10inst0A110` 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 – cf. Hellan 2008, 2010, Hellan and Dakubu 2010.)

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.

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 will instantiate a so-called 'large-to-smart-data' strategy. (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 and Teufel 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. Both are of course currently far beyond what can be realized, partly because of the lack of semantic specifications succinct and consistent enough that they can sustain the kinds of computation envisaged, but they are both methodologically interesting.

## Conclusion - 2

What we have concretely proposed is an analytic formalism which allows us to model the key distinguishing properties of LVCs, namely the situational content of the noun and the degree of unification between noun meaning and verb meaning. 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. The usefulness of such an investigation of LVCs is of course not in doubt, LVCs being essential features of a native speaker's language mastery, so that resources providing them in translation, L2 education, etc., are most valuable. 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.

# 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.
- Brøcker, Karen (2013). “The Danish *går og V* construction in a Semiotic Grammar perspective”. Unpublished paper. Aarhus University.
- Butt, Miriam (2010). “The Light Verb Jungle: Still Hacking Away”. In: *Complex Predicates in Cross-Linguistic Perspective*. Ed. by M. Amberber, M. Harvey, and B. Baker. Cambridge: Cambridge University Press, 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  $\Theta$ -Marking. *Linguistic Inquiry* 19(2):205–232.
- Hellan, L. (2008). Enumerating Verb Constructions Cross-linguistically. In *Proceedings from COLING Workshop on Grammar Engineering Across frameworks*. Manchester. <http://www.aclweb.org/anthology-new/W/W08/#1700>.
- 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

- 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](http://www.typecraft.org/w/images/d/db/1_Introlabels_SLAVOB-final.pdf).
- Jespersen, Otto. 1965. *A Modern English Grammar 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 Samardžić 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.