

# VP idioms in Norwegian: A subconstructional approach

Petter Haugereid

Department of Linguistic, Literary and Aesthetic Studies

University of Bergen

petter.haugereid@lle.uib.no

## Abstract

This paper presents a brief overview of idiomatic expressions in the Norwegian LFG grammar NorGram and shows how the rich lexical information of the LFG grammar can be reused in an HPSG-like grammar with a radically different approach to alternating argument frames. Rather than accounting for idioms by means of special idiom lexical entries, which is the standard approach in LFG and HPSG, a constructional approach is taken where the verbs of the idioms are left underspecified with regard to whether they are idioms or not. A hierarchy of linking types is assumed, which for each piece of evidence provided by the words and rules of the sentence, narrows down the possible frames of the verb to just one.

## 1 Introduction

The Norwegian LFG grammar NorGram (Dyvik, 2000) has 56 VP idioms in the lexicon, distributed over 20 templates. Abstracting away from whether the selected object of the idiom is definite or indefinite, and what kind of argument the selected preposition has (NP, subordinate clause or infinitival clause), we are left with four main kinds of idioms.<sup>1</sup>

The first two kinds of idioms are semantically intransitive, hence they only take one argument, namely the subject. In the first kind of intransitive idioms the main verb selects an object, as

<sup>1</sup>Three idioms (*ta på kreftene* ('tax one's strength'), *sende ord* ('send a message'), and *komme på kant med* ('fall out with')), do not fall into any of the four categories, and they are left out of the present discussion.

shown in (1), and the second kind the main verb selects a PP, as shown in (2).

- (1) Han gikk konkurs.  
he went bankrupt  
*He went bankrupt.*
- (2) De løftet i flokk.  
They lifted in flock  
*They worked together.*

The last two kinds of idioms are semantically transitive, hence they take two arguments. They differ in that in one kind the main verb selects an object and the preposition of a PP, see (3), while in the other the main verb selects a PP and takes an object as an argument, see (4).

- (3) Han la ikke skjul på sin glede.  
he laid not hiding on his joy  
*He didn't hide his joy.*
- (4) Han brakte temaet på bane.  
he brought topic.the on track  
*He brought up the topic.*

A verb that is part of a VP idiom is assigned an idiom frame in the lexicon in addition to the other frames that it appears with. For example the verb *bringe* ('bring') is listed with the following frame:

@(VPIDIOM-PSELOBJ-OBJ bringe på bane)

A lexical entry is allowed to have more than one argument frame by using disjunctions of frames. Disjunctions are expanded into full lexical entries during parsing. This means that a lexical entry with 6 disjunctive argument frames

is computationally equivalent to six lexical entries.

In this paper I will present a new way of representing information about argument frames, including the different kinds of VP idioms presented in this section. The account shifts the burden from the lexicon to a carefully designed hierarchy of linking types. The transfer is achieved by means of *phrasal subconstructions* (see Haugereid and Morey (2012); Haugereid (2012)), which are construction parts that, when put together in a way that conforms with a constraint on the verb, form full constructions. The analysis is implemented in an HPSG-like grammar of Norwegian within the LKB system (Copestake, 2001).

## 2 Treatment of idioms in Sag *et al.* (2003)

In (Sag *et al.*, 2003, 347–355), idioms are assumed to have special lexical entries for the words that constitute them. The idiom *keep tabs on* is analyzed by means of a lexical entry for *keep* (see (5)) with three items on the SUBCAT list; (i) the NP subject, (ii) an idiomatic noun *tabs*, and (iii) a constituent marked by the preposition *on*.

$$(5) \left[ \begin{array}{l} \textit{ptv-lxm} \\ \text{STEM} \langle \textit{keep} \rangle \\ \text{ARG-ST} \langle \text{NP}_i, [\text{FORM } \textit{tabs}], [\text{FORM } \textit{on}] \rangle \\ \text{SEM} \left[ \begin{array}{l} \text{INDEX } s \\ \text{RESTR} \langle \begin{array}{ll} \text{RELN} & \textit{observe} \\ \text{SIT} & s \\ \text{OBSERVER} & i \\ \text{OBSERVED} & j \end{array} \rangle \end{array} \right] \end{array} \right]$$

As (5) shows, the relation of the idiom *keep tabs on* (*observe*) has two arguments, OBSERVER and OBSERVED, and they are linked to the subject of *keep* and the constituent marked by the preposition *on*. Both the idiomatic noun *tabs* and the selected preposition *on* are semantically empty.

Given the degree of detail required in the lexicon, one is forced to assume separate lexical entries for idiomatic verbs. From a semantic point

of view, this is motivated, considering how the meaning of idioms deviates from the compositional meaning. However, there is no morphological evidence indicating that idiomatic verbs should have separate lexical entries. They share the stem with their compositional versions and have the same inflections.

## 3 Analysis

### 3.1 Lexical representation

In addition to the idiom frame given in (4), the verb *bringe* also has a transitive and a ditransitive frame, as shown in (6).

- (6) a. Han brakte maten.  
       he brought food.the  
       *He brought the food.*
- b. Han brakte henne maten.  
       he brought her food.the  
       *He brought her the food.*

Even though we now have three argument frames for the verb *bringe*, we assume only one lexical entry, shown in (7). It has information about the STEM of the lexeme, the HEAD value, the HEAD value of its (potential) arguments, C(ONSTRUCTION)-ARG1, C-ARG2, C-ARG3, and C-ARG4, and its KEYREL.

The four argument features; C-ARG1, C-ARG2, C-ARG3, and C-ARG4 correspond to external subject, (deep) direct object, (deep) indirect object, and oblique object, respectively. Note that there is no linking of the C-ARGS to the semantics. Rather, the linking is done in what we refer to as *phrasal subconstructions*.

$$(7) \left[ \begin{array}{ll} \textit{bringe-v} \\ \text{STEM} & \langle \textit{"bringe"} \rangle \\ \text{HEAD} & \textit{verb} \\ \text{VAL} & \left[ \begin{array}{l} \text{C-ARG1} [\text{HEAD } \textit{noun}] \\ \text{C-ARG2} [\text{HEAD } \textit{noun}] \\ \text{C-ARG3} [\text{HEAD } \textit{noun}] \\ \text{C-ARG4} [\text{HEAD } \textit{compl-noun}] \end{array} \right] \\ \text{KEYREL} & \underline{1} [\text{PRED } \textit{bringe}] \\ \text{RELS} & \langle \underline{1} \rangle \end{array} \right]$$

### 3.2 Phrasal subconstructions

One example of a phrasal subconstruction is the rule that links (external) subjects, *arg1-struct*, illustrated in (8). In this rule, the value of C-ARG1|LINK is switched from *arg1-* in the mother to *arg1+* in the daughter. The C-ARG features that are not mentioned in the representation, as well as the HEAD and KEYREL features are unified with those of the first daughter.<sup>2</sup> At the same time, the argument (the second daughter of the rule) is linked to the ARG1 of the KEYREL.

$$(8) \left[ \begin{array}{l} \textit{arg1-struct} \\ \text{VAL|C-ARG1|LINK } \textit{arg1-} \\ \text{KEYREL } \boxed{1} \left[ \text{ARG1 } \boxed{2} \right] \\ \text{ARGS} \left\langle \left[ \begin{array}{l} \text{VAL|C-ARG1 } \boxed{3} \left[ \begin{array}{l} \text{LINK } \textit{arg1+} \\ \text{INDEX } \boxed{2} \end{array} \right] \\ \text{KEYREL } \boxed{1} \end{array} \right], \boxed{3} \right\rangle \end{array} \right]$$

The grammar also has subconstructions that in the same fashion link (deep) direct objects *arg2-struct*, (deep) indirect objects *arg3-struct*, and oblique objects *arg4-struct*.

The grammar has a rule *vbl-struct* which adds the verb. The verb is selected via the VBL feature, and the VBL value of the verb is transferred to the mother. The rule also unifies its KEYREL value with that of the verb.

$$(9) \left[ \begin{array}{l} \textit{vbl-struct} \\ \text{VBL } \boxed{1} \\ \text{KEYREL } \boxed{2} \\ \text{ARGS} \left\langle \left[ \begin{array}{l} \text{VBL } \boxed{3} \\ \text{KEYREL } \boxed{2} \end{array} \right], \boxed{3} \left[ \begin{array}{l} \textit{verb-word} \\ \text{VBL } \boxed{1} \\ \text{KEYREL } \boxed{2} \end{array} \right] \right\rangle \end{array} \right]$$

### 3.3 Analysis of VP idioms

The analysis of idioms includes phrasal subconstructions for selected prepositions *prepsel-struct* and two subconstructions for idiomatic nouns; *arg2-idiom-struct* and *arg4-idiom-struct*. The *prepsel-struct* subconstruction unifies the FORM value of the selected preposition with the C-ARG4|LINK value, as shown in (10). The features HEAD, VAL, and KEYREL are unified with those of the first daughter (suppressed).

<sup>2</sup>This information is suppressed in the representation in order to save space.

$$(10) \left[ \begin{array}{l} \textit{prepsel-struct} \\ \text{ARGS} \left\langle \left[ \begin{array}{l} \text{V|C-ARG4|LINK } \boxed{8} \\ \text{FORM } \boxed{8} \end{array} \right], \left[ \begin{array}{l} \textit{prep-word} \\ \text{FORM } \boxed{8} \end{array} \right] \right\rangle \end{array} \right]$$

The subconstructions *arg2-idiom-struct* and *arg4-idiom-struct* unifies the FORM value of the selected idiomatic noun with the respective LINK value of the first daughter. This is illustrated for *arg2-idiom-struct* in (11). The C-ARG features not mentioned in the representations, as well as the HEAD and KEYREL features are unified with those of the first daughter. Note that *arg2-idiom-struct* and *arg4-idiom-struct* do not link the idiomatic noun to the KEYREL in the way the *arg1-struct* in (8) does. In this way, the idiomatic nouns do not become semantic arguments.

$$(11) \left[ \begin{array}{l} \textit{arg2-idiom-struct} \\ \text{V|C-ARG2|LINK } \textit{arg2-} \\ \text{ARGS} \left\langle \left[ \begin{array}{l} \text{V|C-ARG2|LINK } \boxed{8} \\ \text{FORM } \boxed{8} \end{array} \right], \left[ \begin{array}{l} \textit{noun-word} \\ \text{FORM } \boxed{8} \end{array} \right] \right\rangle \end{array} \right]$$

The LINK values have two functions. One is to keep track of arguments that are realized. In this way they function like empty/non-empty valence lists in HPSG. The other function is to narrow down what kind of construction the clause has. Each subconstruction that applies in a clause leaves a trace; either a LINK value is switched from negative to positive, or a selected item leaves a mark by unifying its FORM value with the respective LINK value. At the bottom of the tree, all this information is present. This is illustrated in the tree in Figure 1.<sup>3</sup> In the top node *arg4-idiom-struct*, all LINK values are negative, and at the bottom of the tree, in the *START* node, marks from all the subconstructions that have applied can be found.

The *START* sign inherits from the type *uni-link*, which unifies the LINK values of the arguments with the KEYREL|PRED value (see (12)).<sup>4</sup>

<sup>3</sup>The motivation behind the left-branching design is given in Haugereid and Morey (2012).

<sup>4</sup>The unification performed in *uni-link* is left out in the *START* sign in Figure 1, of expository reasons.

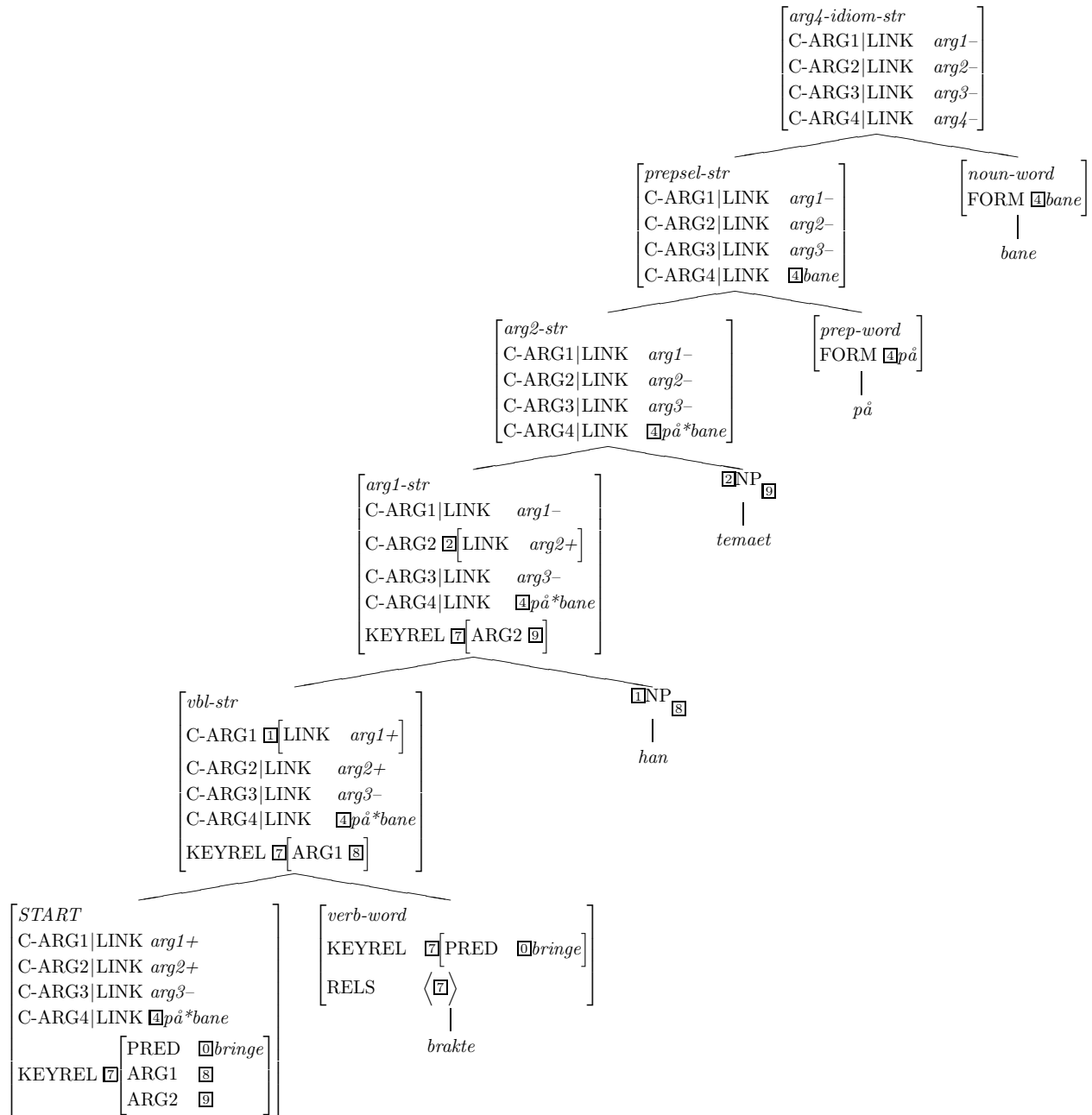


Figure 1: Linking information in the idiom *Brakte han temaet på bane?* (Did he bring up the topic?)

$$(12) \left[ \begin{array}{l} \text{uni-link} \\ \text{VAL } \left[ \begin{array}{l} \text{C-ARG1|LINK } \boxed{1} \\ \text{C-ARG2|LINK } \boxed{1} \\ \text{C-ARG3|LINK } \boxed{1} \\ \text{C-ARG4|LINK } \boxed{1} \end{array} \right] \\ \text{KEYREL|PRED } \boxed{1} \end{array} \right]$$

the KEYREL|PRED value, we get the type  $\text{bringe}^*\text{p}\hat{\text{a}}*\text{bane\_rel}$ . This follows from the hierarchy of linking types showed in Figure 2. The unification of the types  $\text{arg1+}$ ,  $\text{arg2+}$ ,  $\text{arg3-}$ ,  $\text{p}\hat{\text{a}}*\text{bane}$ , and  $\text{bringe}$  gives the type  $\text{bringe}^*\text{p}\hat{\text{a}}*\text{bane\_rel}$ .

When the LINK values of the START sign in the tree in Figure 1 are unified with

The hierarchy in Figure 2 also accounts for the ability of *bringe* to alternate between the

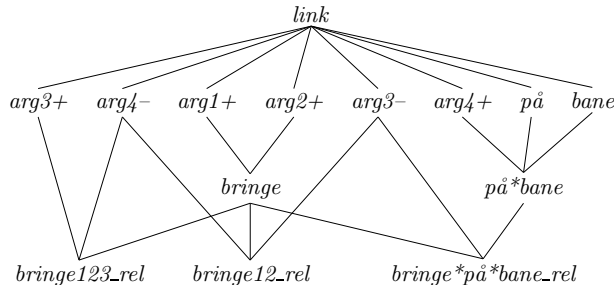


Figure 2: The position of *bringe* in the *link* type hierarchy

three argument frames shown in (4) and (6). Given the three subtypes of *bringe* in the linking type hierarchy, *bringe12\_rel*, *bringe123\_rel*, and *bringe\*på\*bane\_rel*, the verb is allowed to enter the constellations of subconstructions that make up the regular transitive and ditransitive frames as well as the idiom frame.

The MRS (Copestake *et al.*, 2005) produced by the analysis in Figure 1 is given in Figure 13.<sup>5</sup>

$$(13) \left[ \begin{array}{ll} \text{LTOP} & \text{h1} \\ \text{INDEX} & \text{e2}\{\text{pres yes-no-question}\} \\ \text{RELS} & \left. \begin{array}{l} \text{h3:pron\_rel}(x4) \\ \text{h5:pronoun\_q\_rel}(x4,\text{h6},\text{h7}) \\ \text{h8:bringe*på*bane\_rel}(\text{e2},x4,x9) \\ \text{h10:tema\_n\_rel}(x9) \\ \text{h11:def\_q\_rel}(x9,\text{h12},\text{h13}) \end{array} \right\} \\ \text{HCONS} & \{\text{h6 =q h3 h12 =q h10}\} \end{array} \right]$$

The four kinds of idiomatic expression types introduced in Section 1 are accounted for by the following combinations of subconstructions:

1. Intransitive with idiomatic noun ((1)): *vbl-struct*, *arg1-struct*, *arg2-idiom-struct*
2. Intransitive with idiomatic PP ((2)): *vbl-struct*, *arg1-struct*, *prepsel-struct*, *arg4-idiom-struct*
3. Transitive with idiomatic noun ((3)):

<sup>5</sup>The representation presupposes semantically empty versions of the preposition *på* and the noun *bane* as assumed in Sag *et al.* (2003). Alternatively, we could use the regular words. We would then avoid double lexical entries for these words. Their relations *på\_p\_rel* and *bane\_n\_rel* would then be given the same handle as the idiom relation *bringe\*på\*bane\_rel*.

*vbl-struct*, *arg1-struct*, *arg2-idiom-struct*, *prepsel-struct*, *arg4-struct*

4. Transitive with idiomatic PP ((4)): *vbl-struct*, *arg1-struct*, *arg2-struct*, *prepsel-struct*, *arg4-idiom-struct*

## 4 Discussion and future work

The hierarchy of linking types that results from this account is huge, but finite. This kind of hierarchy is interesting in that it reflects what kinds of phrasal subconstructions are needed in order to express all grammaticalized concepts in a given grammar. The grammar presented in this paper, so far only has a small hand-built type hierarchy. The aim is to generate a full hierarchy from the lexicon of the LFG grammar NorGram.

## References

- Copestake, A. (2001). *Implementing Typed Feature Structure Grammars*. CSLI Lecture Notes. Center for the Study of Language and Information, Stanford.
- Copestake, A., Flickinger, D., Pollard, C. J., and Sag, I. A. (2005). Minimal recursion semantics: an introduction. *Research on Language and Computation*, **3**(4), 281–332.
- Dyvik, H. (2000). Nødvendige noder i norsk: Grunntrekk i en leksikalsk-funksjonell beskrivelse av norsk syntaks. In Ø. Andersen, K. Fløttum, and T. Kinn, editors, *Menneske, språk og felleskap*. Novus forlag.
- Haugereid, P. (2012). A grammar design accommodating packed argument frame information on verbs. *International Journal of Asian Language Processing*, **22**(3), 87–106.
- Haugereid, P. and Morey, M. (2012). A left-branching grammar design for incremental parsing. In S. Müller, editor, *Proceedings of the 19th International Conference on Head-Driven Phrase Structure Grammar, Chungnam National University Daejeon*, pages 181–194.
- Sag, I. A., Wasow, T., and Bender, E. M. (2003). *Syntactic Theory: A Formal Introduction*. CSLI Publications, Stanford, 2 edition.