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 constructive 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.\(^1\)

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

(1)  
\begin{align*}
\text{\textbf{Han} gikk konkurs.} \\
\text{he went bankrupt} \\
\text{He went bankrupt.}
\end{align*}

(2)  
\begin{align*}
\text{\textbf{De løftet i flokk.}} \\
\text{They lifted in flock} \\
\text{They worked together.}
\end{align*}

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)  
\begin{align*}
\text{\textbf{Han la ikke skjul på sin glede.}} \\
\text{he laid not hiding on his joy} \\
\text{He didn’t hide his joy.}
\end{align*}

(4)  
\begin{align*}
\text{\textbf{Han brakte temaet på bane.}} \\
\text{he brought topic the on track} \\
\text{He brought up the topic.}
\end{align*}

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 \textit{bringe} (‘bring’) is listed with the following frame:

\begin{align*}
\emptyset\text{(VPIDIOM-PSELOBJ-OBJ bringe på bane)}
\end{align*}

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

---

\(^1\)Three idioms (\textit{ta på kreftene} (‘tax one’s strength’), \textit{sende ord} (‘send a message’), and \textit{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.
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*.

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, (CONSTRUCTION)-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.
3.2 Phrasal subconstructions

One example of a phrasal subconstruction is the rule that links (external) subjects, \textit{arg1-struct}, illustrated in (8). In this rule, the value of \textit{C-ARG1|LINK} is switched from \textit{arg1–} in the mother to \textit{arg1+} in the daughter. The \textit{C-ARG} features that are not mentioned in the representation, as well as the \textit{HEAD} and \textit{KEYREL} features are unified with those of the first daughter.\footnote{This information is suppressed in the representation in order to save space.} At the same time, the argument (the second daughter of the rule) is linked to the \textit{ARG1} of the \textit{KEYREL}.

\begin{equation}
\left[ \begin{array}{c}
\text{arg1-struct} \\
\text{VAL|C-ARG1|LINK arg1–} \\
\text{KEYREL \textcircled{[}ARG1\textcircled{]}} \\
\text{ARGS} \left\langle \left[ \text{VAL|C-ARG1|[LINK arg1+] INDEX \textcircled{[} \textcircled{]} \textcircled{]} \textcircled{]} \right] \right\rangle \\
\end{array} \right)
\end{equation}

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

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

\begin{equation}
\left[ \begin{array}{c}
\text{vbl-struct} \\
\text{VBL \textcircled{[}}} \\
\text{KEYREL \textcircled{[}}} \\
\text{ARGS} \left\langle \left[ \text{VBL \textcircled{[}}} \text{VBL \textcircled{[}}} \text{KEYREL \textcircled{[}}} \text{verb-word \textcircled{[}}} \text{KEYREL \textcircled{[}}} \text{INDEX \textcircled{[}}} \text{FORM \textcircled{[}}} \text{FORM \textcircled{[}}} \text{FORM \textcircled{[}}} \text{FORM \textcircled{[}}} \text{FORM \textcircled{[}}} \right\rangle \\
\end{array} \right)
\end{equation}

3.3 Analysis of VP idioms

The analysis of idioms includes phrasal subconstructions for selected prepositions \textit{prepsel-struct} and two subconstructions for idiomatic nouns; \textit{arg2-idiom-struct} and \textit{arg4-idiom-struct}. The \textit{prepsel-struct} subconstruction unifies the \textit{FORM} value of the selected preposition with the \textit{C-ARG4|LINK} value, as shown in (10). The features \textit{HEAD}, \textit{VAL}, and \textit{KEYREL} are unified with those of the first daughter (suppressed).\footnote{The motivation behind the left-branching design is given in Haugereid and Morey (2012).}

\begin{equation}
\left[ \begin{array}{c}
\text{prepsel-struct} \\
\text{ARGS} \left\langle \left[ \text{V|C-ARG4|LINK \textcircled{[}}} \text{FORM \textcircled{[}}} \right] \right\rangle \\
\end{array} \right)
\end{equation}

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

\begin{equation}
\left[ \begin{array}{c}
\text{arg2-idiom-struct} \\
\text{V|C-ARG2|LINK arg2–} \\
\text{ARGS} \left\langle \left[ \text{V|C-ARG2|LINK \textcircled{[}}} \text{noun-word \textcircled{[}}} \text{FORM \textcircled{[}}} \right] \right\rangle \\
\end{array} \right)
\end{equation}

The \textit{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 \textit{LINK} value is switched from negative to positive, or a selected item leaves a mark by unifying its \textit{FORM} value with the respective \textit{LINK} value. At the bottom of the tree, all this information is present. This is illustrated in the tree in Figure 1.\footnote{The unification performed in \textit{uni-link} is left out in the \textit{START} sign in Figure 1, of expository reasons.} In the top node \textit{arg4-idiom-struct}, all \textit{LINK} values are negative, and at the bottom of the tree, in the \textit{START} node, marks from all the subconstructions that have applied can be found.

The \textit{START} sign inherits from the type \textit{uni-link}, which unifies the \textit{LINK} values of the arguments with the \textit{KEYREL|PRED} value (see (12)).
When the `link` values of the `start` sign in the tree in Figure 1 are unified with the `keyrel|pred` value, we get the type `bringe*på*bane_rel`. This follows from the hierarchy of linking types showed in Figure 2. The unification of the types `arg1+`, `arg2+`, `arg3–`, `på*bane`, and `bringe` gives the type `bringe*på*bane_rel`.

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

Figure 1: Linking information in the idiom *Brakte han temaet på bane?* (Did he bring up the topic?)
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 13.5

\[ \begin{align*}
\text{LTOP} & : h1 \\
\text{INDEX} & : \{ \text{pres yes-no-question} \} \\
\text{RELS} & : \{ \\
& \quad \cdot h3: \text{pron\_rel}(x4) \\
& \quad \cdot h5: \text{pronoun\_q\_rel}(x4,h6,h7) \\
& \quad \cdot h8: \text{bringe\_på\_bane\_rel}(e2,x4,x9) \\
& \quad \cdot h10: \text{tema\_n\_rel}(x9) \\
& \quad \cdot h11: \text{def\_q\_rel}(x9,h12,h13) \\
\text{HCONS} & : \{ h6 = q h3 h12 = q h10 \} \\
\end{align*} \]

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)): \text{vbl-struct, arg1-struct, arg2-idiom-struct} \\
2. Intransitive with idiomatic PP ((2)): \text{vbl-struct, arg1-struct, prepsel-struct, arg4-idiom-struct} \\
3. Transitive with idiomatic noun ((3)): \text{vbl-struct, arg1-struct, arg2-idiom-struct, prepsel-struct, arg4-struct} \\
4. Transitive with idiomatic PP ((4)): \text{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


