

Chapter 3

Multiword expressions in an LFG grammar for Norwegian

Helge Dyvik

University of Bergen

Gyri Smørdal Losnegaard

University of Bergen

Victoria Rosén

University of Bergen

This chapter describes the analysis of multiword expressions in NorGram, an LFG grammar of Norwegian. All multiword expressions need to be accounted for in the lexicon, but in different ways depending on the flexibility of the expression. Each multiword expression is provided with a lexical entry that has a special predicate name incorporating the lexical items that the multiword consists of and that specifies the argument structure of the predicate. In this way, analyses are provided for a wide range of multiword types, including fixed expressions, phrasal verbs, verbal idioms, and others.

1 Introduction

In this chapter we¹ show how multiword expressions (MWEs) are represented in NorGram, a hand-written computational grammar of Norwegian (Dyvik 2000). The grammar is couched in the Lexical-Functional Grammar (LFG) formalism (Bresnan 2001; Dalrymple 2001). It was first developed in the context of the Parallel Grammar Project (ParGram), an international cooperative effort to develop

¹The authors have contributed equally and are listed in alphabetical order.



parallel LFG grammars for a number of languages (Butt et al. 2002). The Xerox Linguistic Environment (XLE) is the platform we use for grammar development and parsing (Maxwell & Kaplan 1993).

NorGram contains about 380 complex syntactic rules, corresponding to a transition network with more than 160,000 states and more than 4.7 million arcs. The lexicon comprises approximately 180,000 lemmas for Norwegian Bokmål and 110,000 lemmas for Norwegian Nynorsk. NorGram uses not only the grammar rules and the lexicon but also templates to efficiently encode linguistic generalizations. As noted in Dalrymple et al. (2004: 207), templates in LFG grammars “can play the same role in capturing linguistic generalizations as hierarchical type systems in theories like HPSG”. Templates are for instance used to express generalizations about subcategorization frames for verbs; there are more than 200 such verbal templates.

NorGram analyzes several types of MWEs, including fixed and flexible expressions. The classification of MWEs according to their relative flexibility was initially proposed for English (Sag et al. 2002; Baldwin & Kim 2010), presupposing that MWEs with the same degree of flexibility may receive the same or similar treatment in NLP systems. The distinction between fixed, semi-fixed and syntactically flexible MWEs may thus be useful also for other languages than English, although the criteria for distinguishing between the classes may vary.

Fixed MWEs are found in most languages with MWEs and in basically every part of speech. These are expressions that are completely invariable, with no morphosyntactic variation or internal modification, such as the adverb *by the way* and the determiner *each and every*. Semi-fixed MWEs, as defined for English, allow some lexical and morphological variation such as limited internal modification and inflection, while the relative word order of the components does not change. Examples are compound nominals (*chicken soup*), proper names, such as *Donald Duck*, and the subset of verbal idioms with fixed word order, such as *shoot the breeze* ‘chat’ and *kick the bucket* ‘die’. Syntactically-flexible expressions display a wider range of flexibility, allowing some or all types of syntactic variation including passivization, relativization and other operations that are not possible in semi-fixed MWEs. All flexible MWEs are verbal. They include verb-particle constructions, light verbs, and the subset of verbal idioms whose word order is less restricted than semi-fixed expressions. Table 1 illustrates how common types of English MWEs distribute over these classes.

The syntactic variation in verbal MWEs in English has given rise to a theory of semantic decomposability (Nunberg et al. 1994) which has led to increased interest in the relation between the syntax and semantics of verbal MWEs. Semantic decomposability is a measure of whether the meaning of the expression

Table 1: Classes of flexibility

Flexibility class	Type	Example MWE
Fixed		<i>by the way</i>
Semi-fixed	compound nominals	<i>chicken soup</i>
	proper names	<i>Donald Duck</i>
	non-decomposable idioms	<i>kick the bucket</i>
Flexible	verb-particle constructions	<i>give up</i>
	light verbs	<i>give a speech</i>
	decomposable idioms	<i>spill the beans</i>

distributes over the MWE components or only relates to the expression as a whole. It may explain why individual parts of an expression may be fronted, topicalized, and relativized, and may also in other ways contribute meaningfully to the information structure of the sentence. On the other hand, semantic non-decomposability blocks compositional interpretations, which again explains why semi-fixed MWEs are not subject to operations that would normally indicate that their components are associated with some independent meaning.

While a distinction between semantically decomposable and nondecomposable verbal idioms may also hold for Norwegian, the correlation between syntactic flexibility and semantic decomposability seems less conspicuous than for English. In particular, Norwegian has subject-verb inversion in interrogative main clauses, so that the word order will vary in MWEs that are otherwise highly restricted. Most verbal idioms may also undergo at least some modification (e.g., impersonal passives). Furthermore, the mechanisms for representing restrictions and variation in NorGram are technically the same for semi-fixed and flexible MWEs. Since no distinction is reflected in the way verbal MWEs are represented in the lexicon and grammar, all such MWEs are considered flexible, and MWEs with similar morphosyntactic properties are accounted for with templates which are in effect mini-grammars for subsets of MWEs.

With respect to subtypes of MWEs, the types of MWEs analyzed by NorGram more or less correspond to the types in Table 1, with a few exceptions. As in many other Germanic languages, compound nominals in Norwegian form single graphical words. These are thus not considered multiword expressions. In addition to prepositional verbs, NorGram analyzes nouns and adjectives with selected prepositions as MWEs. Expressions that are completely regular on the morphological

and syntactic levels, such as light verb constructions, are analyzed compositionally by the grammar and are not represented in the lexicon as MWEs. A special case is complex numerals such as *hundre og to* ‘one hundred two’ and *to og nitti* ‘ninety two’, which may also be considered a subtype of MWE. The particular syntax and semantics of such expressions is accounted for with a special set of lexical entries and syntactic rules.

NorGramBank, a large parsebank for Norwegian, has been created by parsing a corpus with NorGram (Dyvik et al. 2016). Because of lexical and syntactic ambiguity, parsing with NorGram often results in many analyses for each sentence, and efficient disambiguation is therefore necessary. The INESS project² has developed a treebanking infrastructure for parsing, disambiguating, storing, and searching the texts in NorGramBank (Rosén et al. 2012). The parsebank currently consists of about 60 million words of analyzed text, of which sentences covering 350,000 words have been manually disambiguated by computer-generated discriminants (Rosén et al. 2007). The remainder of the corpus has been stochastically disambiguated. INESS Search is a tool for searching in LFG and other treebanks in the treebanking infrastructure (Meurer 2012). MWEs are analyzed by NorGram in such a way that the different types may be searched for.

The original lexical resource used for the NorGram lexicon, NorKompLeks, contained a small number of fixed expressions (Nordgård 2000). The main design of the treatment of MWEs in NorGram was developed during ParGram (Butt et al. 2002) and especially during the LOGON machine translation project (Lønning et al. 2004). A large number of MWEs have been added to NorGram’s lexicon during the construction of NorGramBank. When disambiguators discovered MWEs that did not receive an analysis or that received an incorrect analysis, they constructed new lexical entries or edited existing lexical entries as needed in order to cover the MWEs (Losnegaard et al. 2012; Rosén et al. 2016).

This chapter is organized as follows. In Section 2 an overview of the basics of LFG is given, showing how constructions without MWEs are analyzed in NorGram as a background for the treatment of MWEs in the following sections. Section 3 illustrates the NorGram analysis of MWEs, both fixed expressions and flexible expressions such as phrasal verbs, verbal idioms, and nonverbal flexible expressions. Section 4 shows how various syntactic modifications are handled, including intervening words, long-distance dependencies and passive alternations. Section 5 discusses numerous complex complementation patterns that are covered by NorGram for Norwegian MWEs. Section 6 presents our conclusions.

²<http://clarino.uib.no/iness>

2 Syntactic analysis in LFG

LFG analyses have two distinct levels of syntactic representation: constituent structure (c-structure) and functional structure (f-structure). The c-structure is a phrase structure tree that represents precedence and dominance relations. The f-structure is an attribute-value matrix with information about grammatical functions such as subject and object and grammatical features such as tense, gender and number. An example of a NorGram analysis of the sentence in (1) is given in Figure 1.³

- (1) Hun tenkte på buss-en.
 she thought on bus-DEF.SG
 ‘She was thinking (while) on the bus./She thought about the bus.’

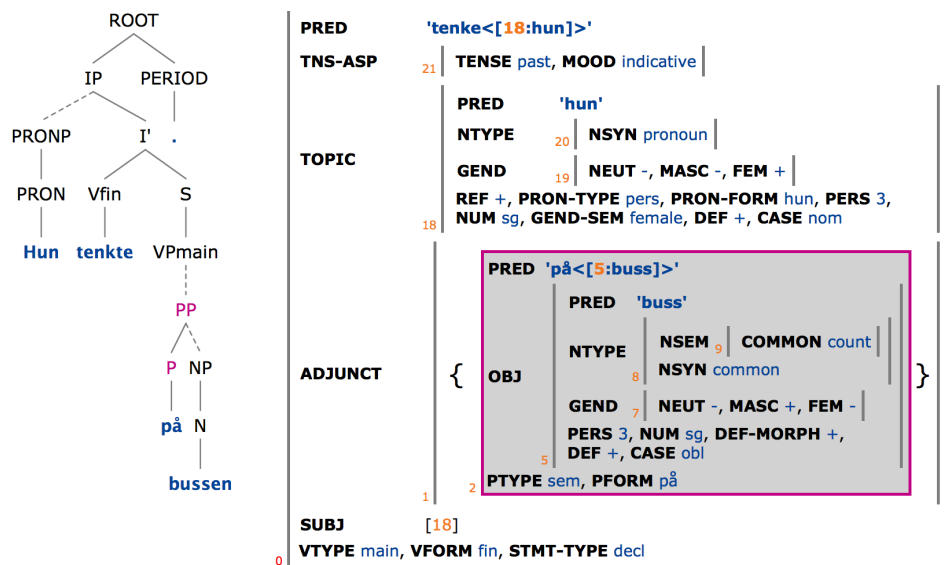


Figure 1: C- and f-structure for *Hun tenkte på bussen*.

This sentence is ambiguous, as shown by the two idiomatic translations. The analysis in Figure 1 concerns the first translation, where the prepositional phrase

³In this example the morphological structure of the word form *bussen* is indicated since it is relevant for the analysis being discussed. Otherwise, we simplify the glossing by omitting morpheme-by-morpheme analysis and using two English words to render one Norwegian word when necessary.

på bussen ‘on the bus’ is an adjunct (adverbial). The second reading, where *tenke på* ‘think about’ is a phrasal verb, will be treated in Section 3.2.

The phrase structure rules and lexicon of an LFG grammar assign the c-structure. NorGram uses a version of X'-syntax that is inspired by Bresnan (2001), with some adjustments which depart from strictly binary branching structures.

The f-structure is projected from the c-structure by the functional description (f-description), which describes correspondences between the two levels. One such correspondence is illustrated in Figure 1 by the highlighting of the PP node and the corresponding partial f-structure. The phrase structure rules that assign this f-structure are given in (2) and (3). The rule daughters are listed vertically after the horizontal arrow, with each node's functional annotations following after a colon.⁴

$$(2) \quad \text{PP} \rightarrow \text{P: } \uparrow = \downarrow \\ \quad \quad \quad \text{NP: } (\uparrow \text{ OBJ}) = \downarrow$$

$$(3) \quad \text{NP} \rightarrow \text{N: } \uparrow = \downarrow$$

The annotations on the rule daughters describe the associated f-structures. In the equations, \uparrow refers to the f-structure of the mother node (the category on the left-hand side of the rule), while \downarrow refers to the f-structure of the daughter node (the category carrying the annotation on the right-hand side of the rule). Thus the equation $\uparrow = \downarrow$ annotated to a rule daughter means that the daughter node and its mother node will project the same f-structure. The equation $(\uparrow \text{ OBJ}) = \downarrow$ on the NP node in (2) specifies that the f-structure of the mother node (PP) has an object (OBJ) which is the f-structure of the daughter node (NP). In this way the highlighted f-structure with the index “2” at its lower left corner in Figure 1 is projected from the PP node. Both the PP node and the P node are highlighted in the c-structure since they both project this same f-structure.

The annotations on the phrase structure rules account for only part of the information in the f-structure. Other information comes from the word forms in the terminal nodes of the tree. For instance, the lexical and morphological information for the word *bussen* contributes all the equations in (4). These equations are part of the f-description for the f-structure that is the value of the OBJ attribute (with the index “5”) in Figure 1.

⁴The examples of rules, lexical entries, and templates in the following are simplified for the purpose of exposition. Neither the format nor the content is exactly the same as in NorGram.

- (4) (\uparrow PRED) = 'buss'
 (\uparrow NTYPE NSEM COMMON) = count
 (\uparrow NTYPE NSYN) = common
 (\uparrow GEND NEUT) = -
 (\uparrow GEND MASC) = +
 (\uparrow GEND FEM) = -
 (\uparrow PERS) = 3
 (\uparrow NUM) = sg
 (\uparrow DEF-MORPH) = +
 (\uparrow DEF) = +

The first equation, which assigns the PRED(icate) value 'buss', is specific to this noun, but the others are common to many other words. Some of the equations come from features assigned to the word form *bussen* by the morphological analyzer run prior to parsing; these features are +Noun, +Sg, +Def and +Masc, and they will appear in the string presented to the syntactic parser. Other equations come from the lexical entry for the noun *buss*. Both the features and the noun must have entries in the lexicon; these are shown in (5–9). Each lexical entry specifies a lexical category; SUFF (for suffix) is the category for morphological features.

- (5) +Noun SUFF (\uparrow PERS) = 3
 (6) +Sg SUFF (\uparrow NUM) = sg
 (7) +Def SUFF @DEF
 (8) +Masc SUFF @MASC
 (9) buss N @(COUNTNOUN buss)

The equations in the first two entries each contribute one attribute-value pair to the f-structure. Entries (7–9) contain template invocations rather than equations. The @-sign indicates a call to a template, while DEF, MASC and COUNTNOUN are names of templates. A template is an f-description, a collection of equations which it is convenient to refer to by a name rather than listing all the equations. Templates can be used in different places in the grammar and lexicon, and template definitions may refer to other templates.

The definition of the template named DEF is shown in (10). All nouns inflected in the definite form will carry these two equations, so it can be convenient to refer to them together.

- (10) DEF =
 (↑ DEF-MORPH) = +
 (↑ DEF) = +

Norwegian has a complicated system of gender agreement because of some nouns that may have either masculine or feminine agreement, and because adjectives and determiners may be unspecified for certain gender distinctions. To account for this, each noun must receive a plus or minus value for each of the three genders. The equations needed for specifying masculine gender are included in the template in (11). These equations do not simply describe attribute-value pairs; they describe paths through the f-structure. The equation (↑ GEND MASC) = + states that the f-structure has an attribute GEND which has as its value a subsidiary f-structure which in its turn has an attribute MASC with the value +.

- (11) MASC =
 (↑ GEND MASC) = +
 (↑ GEND FEM) = -
 (↑ GEND NEUT) = -

Like the template MASC, the template COUNTNOUN also describes paths through the f-structure. The NTYPE NSYN features distinguish between common nouns, proper nouns, pronouns, etc. while the NTYPE NSEM COMMON features distinguish between count nouns, mass nouns, etc. All nouns must contribute a PRED feature to the f-structure, but the PRED feature itself will differ from noun to noun. The template in (12) is parameterized; the parameter P will be substituted by the argument supplied in the invocation of the template, for example the word *buss* in (9).

- (12) COUNTNOUN (P) =
 (↑ PRED) = P
 (↑ NTYPE NSEM COMMON) = count
 (↑ NTYPE NSYN) = common

The value of a PRED attribute is a semantic form. A semantic form is always enclosed in single quotation marks, indicating that the value is unique, which means that it cannot be unified even with an identical-looking value of some other attribute. For some words the semantic form includes not only the word itself, but also a syntactic argument list. This is the case for two of the words in *hun tenker på bussen* (in the interpretation being considered in this section). The verb has the semantic form 'tenke<[SUBJ]>', meaning that the verb is intransitive

and subcategorizes only for a subject, and the preposition has the semantic form ‘på⟨[OBJ]⟩’, indicating that it requires an object.

Verbs can of course subcategorize for several arguments. For example, the verb *slå* ‘hit’ has the semantic form ‘slå⟨[SUBJ,OBJ]⟩’ since it requires a subject and an object. The completeness requirement for f-structures stipulates that each of the syntactic functions mentioned in the semantic form of a PRED feature must occur on the same level of f-structure as that PRED. There is also a coherence requirement to the effect that subcategorizable syntactic functions may only occur on the same level of f-structure as a PRED feature if they are mentioned in its semantic form. The argument lists in semantic forms are thus crucial for determining grammaticality. The semantic forms “govern the process of semantic interpretation” (Kaplan & Bresnan 1982: 177).

3 Implementing MWEs

The crucial challenge of representing MWEs is that they defy normal compositional analysis. The LFG solution that is implemented in NorGram is to assign to each MWE a special lexical entry that has its own PRED value and thus its own argument structure. Each MWE has a semantic form with a special predicate name and a list of any syntactic arguments that this predicate requires. This will be shown in detail for the various types of MWEs in the following.

3.1 Fixed expressions

Fixed expressions, such as *ad hoc*, *déjà vu*, and *vice versa*, are those that do not vary with respect to inflection and that do not admit any internal modification. They are also called inflexible expressions or “words with spaces”. Fixed expressions are the simplest MWEs to implement; they are entered into the NorGram lexicon as single graphical words containing white space, so they are literally treated as words with spaces.

- (13) Hun likte **i bunn og grunn** ikke New York **i det hele tatt**.
 she liked in bottom and ground not New York in the whole taken
 ‘She basically didn’t like New York at all.’

The sentence in (13) contains three such expressions: *i bunn og grunn*, *New York*, and *i det hele tatt*.⁵ The c-structure of (13) is shown in Figure 2. The simplified

⁵In this and subsequent examples the lexically fixed words making up the MWE are highlighted with boldface.

f-structure is shown in Figure 3; this is the “PREDS only” view of f-structure where feature paths that do not end in PRED values are suppressed. The three expressions belong to different parts of speech: ADVcmt (commitment adverb), PROP (proper noun), and ADVs (sentence adverb). The adverbs have the function ADJUNCT in the f-structure while the proper noun functions as the OBJ. There are numerous fixed expressions in most parts of speech in Norwegian.

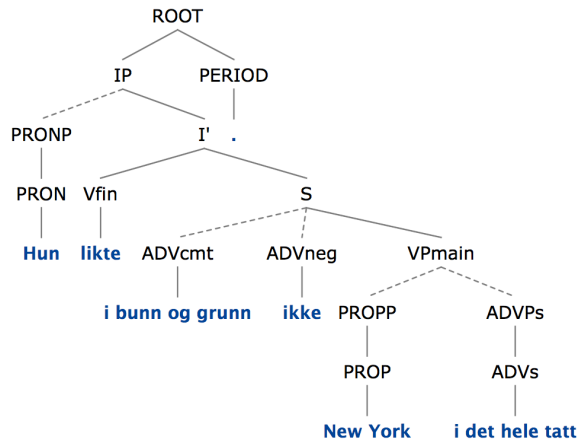


Figure 2: C-structure for example (13)



Figure 3: F-structure for example (13)

3.2 Basic properties of flexible MWEs and method of analysis

Flexible expressions may exhibit a great deal of syntactic variation, but in some respects they are inherently fixed or restricted. One of the characterizing features of MWEs is that they are lexically fixed, meaning that they consist of at least two words that cannot be substituted with near-synonyms or semantically related words without the expression losing its idiomatic meaning. The verbal idiom *komme på kant med* in (14) has four such fixed lexical words.

- (14) Hun var ikke villig til å **komme på kant med** det gode selskap.
 she was not willing to to come on edge with the good company
 ‘She was not willing to fall out with the in-crowd.’

Flexible MWEs are often also morphosyntactically restricted, with constraints on grammatical features, on the modification of component words, and on specifiers such as quantifiers and determiners. For instance, the noun *kant* in (14) can only be in the singular indefinite form, and it does not admit any specifiers or modifiers. The PP *på kant med*, however, does admit modifiers; in (15) the modifier *helt* ‘completely’ has scope over the entire expression. In NorGram, no distinction is currently made in the representation of internal and (semantically) external modification of MWEs.

- (15) Hun **kom helt på kant med** det gode selskap.
 she came fully on edge with the good society
 ‘She completely fell out with the in-crowd.’

The mechanisms for representing lexical and morphological restrictions in flexible MWEs are the same as the ones used for regular constructions. As described in Section 2, simplex words are assigned predicate values through equations in the lexical entry, as in the entry for the simplex lexeme *buss* in example (4), which has the predicate assignment equation (\uparrow PRED) = ‘buss’. For words that subcategorize for other elements, such as verbs, this is done through the assignment of a predicate-argument structure (or subcategorization frame). For instance, the intransitive verb *klage* ‘complain’ is assigned a frame through the template call @ (V-SUBJ *klage*) in the lexical entry, invoking the template V-SUBJ. Part of this template is shown in (16).

- (16) V-SUBJ (P) =
 (\uparrow PRED) = ‘P < (\uparrow SUBJ)’

The predicate value of the verb is parameterized and listed together with its arguments in the subcategorization frame, which includes everything between quotation marks in (16). When the template is invoked, the lemma form *klage* in the template call replaces the parameter P. The equation on the second line assigns one argument, the subject, to P, yielding the predicate-argument structure ‘*klage* < (\uparrow SUBJ)’ as the PRED value for the intransitive reading of this verb.

Lexical fixedness in flexible MWEs is handled through lexical selection in the entry of the subcategorizing word. In addition to its usual intransitive reading, the verb *klage* ‘complain’ is the syntactic head of the VP idiom *klage sin nød* ‘pour out one’s troubles’, where it subcategorizes for the object noun *nød* ‘need’.

- (17) Og hun klaget sin nød til alle som ville høre.
 and she complained her need to all who wanted hear
 ‘And she poured out her troubles to everyone who wanted to listen.’

Lexical entries for VP idioms are listed as alternative subcategorization frames under the entry of the verb. As in the case of simplex verbs, templates assign predicate-argument structures and other relevant features. The word that subcategorizes for the other parts of the MWE lists the predicate values of the selected arguments together with its own predicate value in the template invocation. The template call in (18) shows that the verb *klage* selects the noun *nød*.

- (18) @(VPIDIOM-DEFOBJ klage nød)

- (19) (↑ PRED) = '%FN<(↑ SUBJ)>(↑ OBJ)'

The predicate values of the fixed MWE components, i.e. the verb and its selected complements, are merged to form one single idiom predicate which is substituted for the relevant parameter in the predicate-argument structure. In one of the equations in the template, the predicate assignment in (19), the parameter %FN is replaced by the predicate name *textsflage#nød*, where we use the symbol “#” to signal idiomatic combinations of this kind. Only the free arguments of verbal MWEs are specified as semantic arguments to the verb. The subcategorization frame ‘*klage#nød*<(↑ SUBJ)>(↑ OBJ)’ lists the semantic argument, in this case SUBJ, inside the angled brackets, while the selected argument OBJ is placed outside the brackets. The parameter %FN and the construction of predicate names such as *klage#nød* are accounted for in Section 3.3.1.

Constraints on grammatical features are specified with constraining equations and existential constraints in the entries or templates. A constraining equation is an equation with a “c” attached to the equal sign. This means that the equation does not actually assign the specified value to the attribute in the f-structure; instead it requires that this value has been assigned to the attribute somewhere else. The restriction that the object *sin nød* in (17) must be definite is specified with the equation in (20). The constraint in (21) is an existential constraint which simply provides a path of attributes without assigning a particular value. The interpretation is that this path of attributes must have some value in the f-structure, thus ensuring that there is a possessive.

- (20) (↑ OBJ DEF) = c +.

(21) $(\uparrow \text{OBJ SPEC POSS POSS-TYPE})$

(22) $\sim(\uparrow \text{OBJ SPEC})$

The selection of grammatical words and modifiers is handled in a slightly different way from the selection of syntactic heads. If a determiner is selected or otherwise restricted, this is specified with a constraint requiring that the type or form of the determiner must match the specification. The existential constraint in (21) ensures that a possessive will specify *nød*. If no determiner is possible in an idiom, this is specified with a negative constraint, as in (22).

Lexical constraints on modifiers are represented in the same way as grammatical constraints, using equations. Some nouns do not admit modification at all, such as *kant* in (14). Others may require that the choice of modifier is restricted to a specific predicate or set of predicates, such as *øye* ‘eye’ in the VP idiom *ha et godt øye til* ‘have eyes for’ in (23), where the only possible modifier is the adjective *god* ‘good’.

(23) Det kan være han **har et godt øye til** deg.
 it can be he has a good eye to you
 ‘He might have eyes for you.’

When a modifier is lexically restricted, a constraint equation is used to specify the possible modifier predicate(s). In the entry for *ha et godt øye til*, the equation in (24) ensures that the modifier (ADJUNCT) of the selected object (the noun *øye*) has the PRED value *god*.

(24) $(\uparrow \text{OBJ ADJUNCT PRED}) = c \text{ god}$

The treatment of lexical restrictions in VP idioms in NorGram thus depends on the function of the component word within the MWE. While syntactic heads are subcategorized for by the verb, dependents are specified using constraint equations.

3.3 Phrasal verbs

Phrasal verbs are MWEs consisting of a verb and an adverb, preposition or other word that together have a meaning that is in some way idiosyncratic. It is common to distinguish between two main classes of phrasal verbs, prepositional verbs and verb-particle constructions. We present these two types in Sections 3.3.1 and 3.3.2, respectively. There are also constructions where both prepositions and particles occur; these are presented in Section 3.3.3.

3.3.1 Prepositional verbs

In Section 2 the sentence in (1) *Hun tenkte på bussen* was shown to have two readings. When the prepositional phrase functions as an adjunct, the analysis shown in Figure 1 obtains. When the preposition is selected by the verb, the verb and the preposition constitute an MWE, as indicated in (25), where these words are boldfaced. The analysis corresponding to this reading is shown in Figure 4.

- (25) Hun **tenkte på** bussen.
 she thought on the bus
 ‘She thought about the bus.’

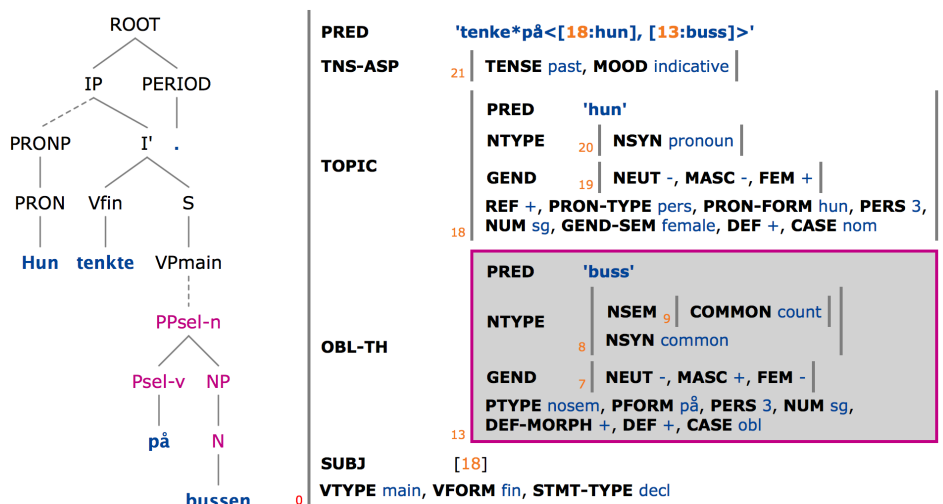


Figure 4: C- and f-structure for example (25)

In the c-structure *på bussen* forms a prepositional phrase *PPsel-n*, marked as selected by *sel-n* in the node label. This analysis captures the fact that the selected preposition *på* can only occur before the object, and that the preposition and its complement behave as one constituent with respect to movement, as in the topicalized version *På bussen tenkte hun ofte* ‘The bus she was often thinking of’. The preposition does not provide its own predicate in the f-structure, but is analyzed as incorporated in the predicate expressed by the verb to form the predicate name *tenke*på*. In predicate names the symbol “*” is used to signal such combinations of a lexical predicate with a selected particle or preposition. The complement of the preposition, *bussen*, fills the function *OBL-TH* – oblique-theta – as an argument of this predicate, i.e., an oblique argument expressing a theta role.

The lexical entry for *tenke* is associated with the relevant frame through an invocation of the template describing this class of constructions. The relevant part of the lexical entry for *tenke* is shown in (26).

- (26) *tenke* V { [...]

 | @(V-SUBJ-POBJ *tenke* på)

 | [...] }

The invocation of the template V-SUBJ-POBJ has two parameters, the predicate name for the verb *tenke* and the form of the selected preposition *på*. In (27) part of the template is shown (other parts of this template for handling passive and other modifications are discussed in Section 4).

- (27) V-SUBJ-POBJ (P prp) =

 @(CONCAT P '* prp %FN)

 (↑ PRED) = '%FN<(↑ SUBJ)(↑ OBL-TH)>'

 (↑ OBL-TH CHECK P-SELFFORM) = prp

The template invokes another template CONCAT, which concatenates the predicate name P and the preposition form prp as the value of the variable %FN. In this example the result is the predicate name *tenke*på*, which is then included in the value of PRED. The last line assigns the value of prp (*på* in the example) as the value of the attribute P-SELFFORM under the OBL-TH argument. This feature is checked by the syntactic rule which introduces the selected PP, ensuring that only the preposition selected by the verb is accepted.

3.3.2 Verb-particle constructions

Verb-particle constructions consist of a verb and a selected particle in the form of an adverb or an intransitively used preposition; in NorGram such elements are classified as PRT in the c-structure. The verb and the particle express an idiosyncratic meaning. As in English, verb-particle constructions in Norwegian can have the particle either before or after an object, and obligatorily after if the object is pronominal; cf. Baldwin & Kim (2010: 276). The analysis is illustrated in Figures 5 and 6 for the sentence in (28).

- (28) Han **skrev opp** nummeret.
 he wrote up the number
 ‘He wrote down the number.’

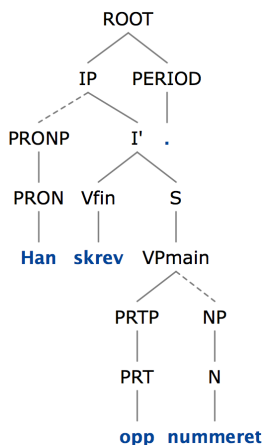


Figure 5: C-structure for example (28)

PRED	'skrive*opp<[13:han], [11:nummer]>'
TNS-ASP	17 TENSE past, MOOD indicative
TOPIC	PRED 'han' NTYPE 15 NSYN pronoun GEND 14 NEUT -, MASC +, FEM - 13 REF +, PRON-TYPE pers, PRON-FORM han, PERS 3, NUM sg, GEND-SEM male, DEF +, CASE nom
OBJ	PRED 'nummer' NTYPE NSEM 9 COMMON count 8 NSYN common GEND 7 NEUT +, MASC -, FEM - 11 PERS 3, NUM sg, DEF-MORPH +, DEF +, CASE obl
SUBJ	[13]
0	VTYPE main, VFORM fin, STMT-TYPE decl, PRT-FORM opp

Figure 6: F-structure for example (28)

In the c-structure the particle PRT is a separate constituent which may also occur after the NP under VPmain. In the f-structure the verb and the particle are analyzed as forming one predicate *skrive*opp*, and the particle also provides a value to the feature PRT-FORM.

As in the case of selected prepositions, the lexical entry for *skrive* is associated with the relevant frame through an invocation of the template describing this class of constructions. Part of the lexical entry for *skrive* is shown in (29).

- (29) *skrive* V { [...]
 | @(V-SUBJ-PRT-OBJ *skrive opp*)
 | [...] }

Part of the invoked template V-SUBJ-PRT-OBJ is shown in (30).

- (30) V-SUBJ-PRT-OBJ (P prt) =
 @(CONCAT P '* prt %FN)
 (↑ PRED) = '%FN<(↑ SUBJ)(↑ OBJ)>'
 (↑ CHECK PRT-VERB) = +
 (↑ PRT-FORM) = c prt

The CONCAT template functions as in the template (27), yielding the predicate name *skrive*opp* as the value of PRED. The second last line assigns the value “+” to the path CHECK PRT-VERB, a feature which is checked by the syntactic rule introducing the particle PRT; see the VPmain rule in (43) below. The last line is a constraining equation⁶ which checks that the value of the feature PRT-FORM, which is introduced in the sentence by the particle, is the value of prt, i.e. *opp* in the template invocation in (29).

3.3.3 Verb-particle constructions with selected prepositions

The preceding sections have shown how prepositional verbs and verb-particle constructions are analyzed. Phrasal verbs also allow both selected prepositions and particles in the same MWE. An example involving both, in addition to a reflexive object, is provided in the treebank example in (31). The analysis of (31) is shown in Figures 7 and 8.

- (31) Vi har et så enormt stort område å sette oss inn i.
 we have a such enormously large area to set us in into
 ‘We have such an enormously large area to immerse ourselves in.’

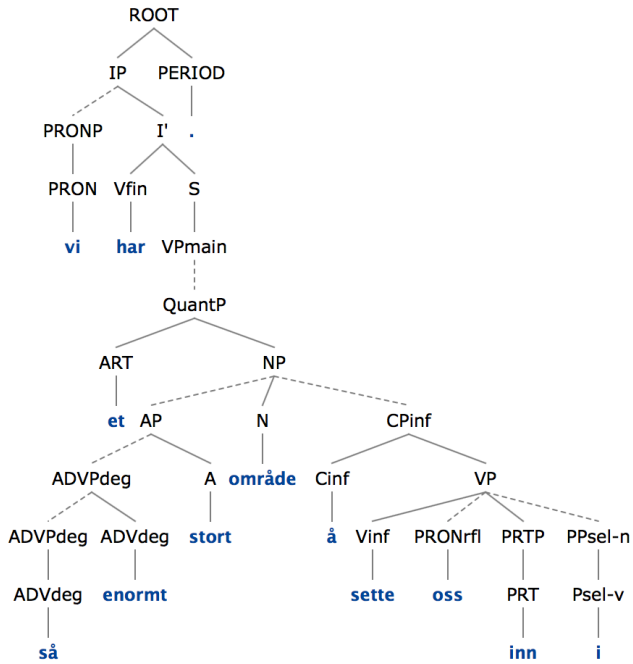


Figure 7: The c-structure of sentence (31)

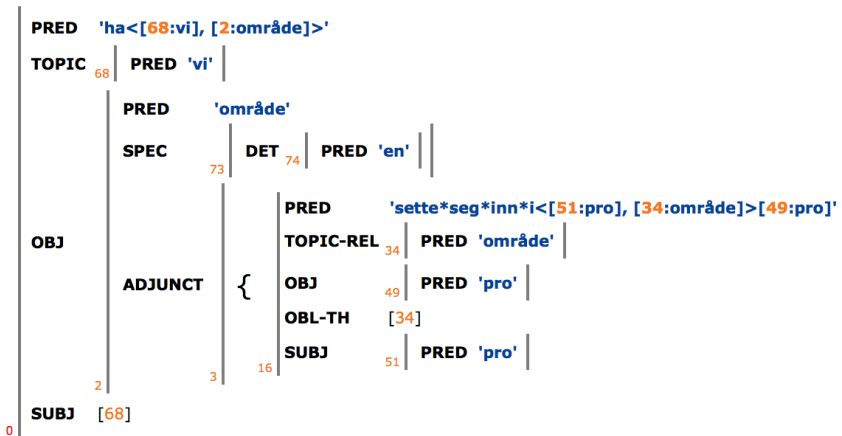


Figure 8: Part of the simplified f-structure of sentence (31)

In this example the MWE *sette oss inn i* occurs in an infinitival relative (CPinf) in an NP with the head *område* ‘area’. In the f-structure the infinitival relative occurs as a member of the set of adjuncts to the predicate *område*, also occurring as the second argument of *sette*seg*inn*i* as its relativized argument (see Section 4.2 for the analysis of long-distance dependencies like relativization and topicalization). The template invoked by the verb *sette*, V-SUBJ-OBJrefl-PRT-POBJ, provides an analysis along the lines of the templates in (27) and (30).

3.4 Verbal idioms

A VP idiom is a flexible MWE in which at least one predicate-bearing lexeme (such as a noun or an adjective) is selected, with possible restrictions as to number, definiteness or other morphological properties applying. VP idioms are handled by a specific set of templates. For example, an idiom like *holde øye med* ‘keep an eye on’ is analyzed by means of a lexical template covering idioms consisting of a selected indefinite object plus a selected prepositional phrase. The treebank sentence in (32) is analyzed as shown in Figures 9 and 10.

- (32) Samtidig **holdt** han umerkelig et skarpt **øye med** jentungen.
 simultaneously kept he unnoticeably a sharp eye with the girl child
 ‘At the same time he furtively kept a close eye on the girl.’

The analysis of the selected prepositional phrase *med jentungen* is as described in Section 3.3.1 for example (25). The selected lexeme in (32) is *øye*. In the f-structure in Figure 10 the idiomatic meaning is represented by incorporating *øye* in the predicate name, deriving the predicate name *holde#øye*med*. The phrase *et skarpt øye* fills the function of OBJ, but is not analyzed as a semantic argument of the sentence predicate, which appears from its position outside the angled brackets <...> surrounding the argument list. This position signals that the constituent is syntactically subcategorized for without being a semantic argument.

The lexical entry for *holde* is associated with the VP idiom through an invocation of the idiom template describing the relevant class of idioms. Part of the lexical entry for *holde* is shown in (33). The template invocation has three parameters, the predicate name for the verb *holde*, the predicate of the selected noun *øye*, and the form of the selected preposition *med*. Part of the template is shown in (34); the full template is discussed in Section 4.3.

⁶This concept is explained in connection with example (20) above.

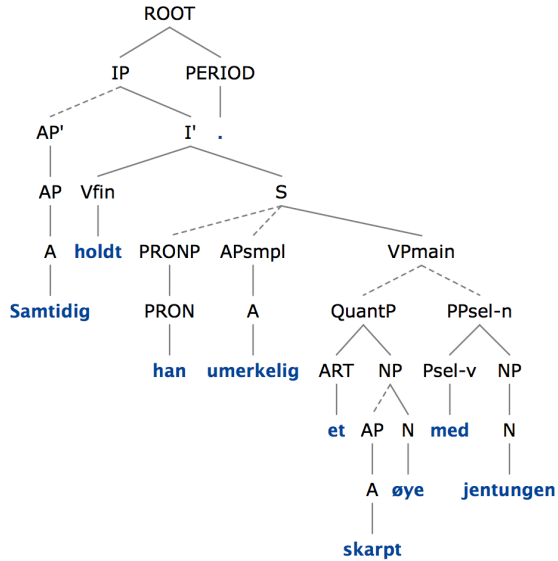


Figure 9: The c-structure of sentence (32)

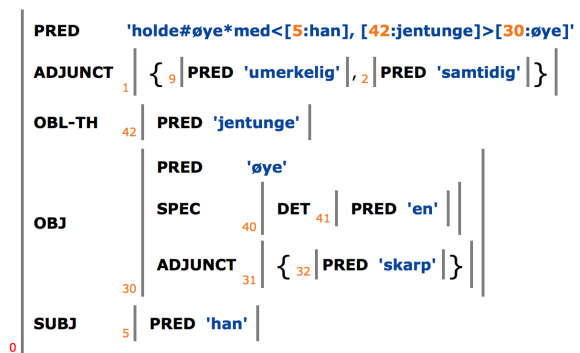


Figure 10: The simplified f-structure of sentence (32)

(33) holde V { [...]
 | @(VPIDIOM-INDEF OBJ-POBJ holde øye med)
 (↑ OBJ NUM) = c sg
 | [...] }

(34) VPIDIOM-INDEF OBJ-POBJ (P OP prp) =
 @(CONCAT P '# OP * prp %FN)
 (↑ PRED) = '%FN<(↑ SUBJ)<(↑ OBL-TH)>(↑ OBJ)
 (↑ OBL-TH CHECK P-SELF FORM) = prp
 (↑ OBJ PRED FN) = c OP
 ~(↑ OBJ DEF) = +

In addition to the template call, the lexical entry in (33) also specifies that the selected object should be singular. It is a matter of choice whether such information should be included in the individual lexical entry or give rise to a distinction between more fine-grained templates. In the template definition in (34), OP (object predicate) is the variable for the selected noun predicate and prp for the selected preposition. As in the case of the template in (30), the template invokes the CONCAT template which builds the predicate name. The second last equation requires the object to have the value of OP as its predicate (in this case 'øye'), and the final equation requires the object not to be definite. As for the equation mentioning P-SELF FORM, see the explanation of the template in (27).

3.5 Nonverbal flexible expressions

3.5.1 Nouns with selected prepositions

Nouns may also form MWEs by selecting prepositional phrases as their arguments. For example, the noun *ansvar* 'responsibility' may select the preposition *for* 'for', which can take a nominal phrase, an infinitival, or a nominal subclause as complement, as in the treebank examples in (35–37).

(35) Hadde jeg **ansvar** **for** gutten?
 had I responsibility for the boy
 'Did I have responsibility for the boy?'

(36) Han fikk **ansvar** **for** å overta søket.
 he got responsibility for to take over the search
 'He got the responsibility for taking over the search.'

- (37) Jeg kan ikke ta ansvar for at det ble lekket.
 I can not take responsibility for that it became leaked
 ‘I cannot take responsibility for its having been leaked.’

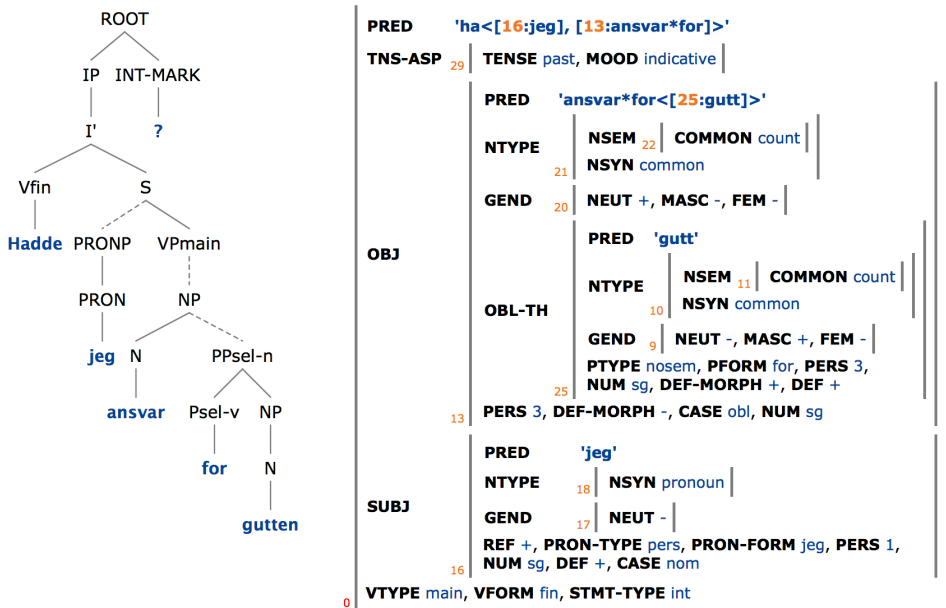


Figure 11: C- and f-structure for the sentence (35)

Example (35) is analyzed as in Figure 11. As in the case of prepositional verbs, the selected preposition does not contribute a PRED of its own, but is analyzed as forming a single predicate *ansvar*for* with the noun, taking the complement of the preposition as an argument with the function OBL-TH (an oblique argument expressing a theta role). With an infinitival or a clausal complement the syntactic function is COMP. The lexical entry for *ansvar* in (38) invokes three alternative templates for the three possible kinds of complements, in addition to its basic template as a mass noun.

- (38) ansvar N { @(MASSNOUN ansvar)
 | @(N-POBJ ansvar for)
 | @(N-PINFCOMP ansvar for)
 | @(N-PCOMP ansvar for) }

3.5.2 Adjectives with selected prepositions

Similarly, adjectives may select prepositional phrases as complements, for instance *flink til* ‘clever at’, as in the treebank example in (39).

- (39) Hva er egentlig du flink til?
 what are after all you clever at
 ‘What are you clever at, after all?’

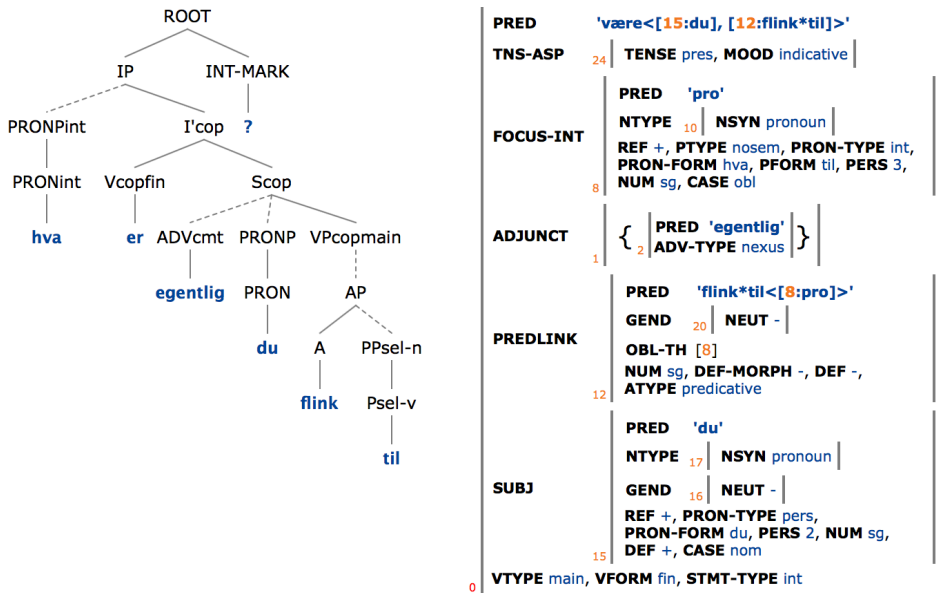


Figure 12: C- and f-structure for the sentence (39)

Example (39) is analyzed as in Figure 12. In this example the complement of the selected preposition has been questioned and occurs in the f-structure as the value of FOCUS-INT, i.e., interrogative focus. The predicative complement (PREDLINK) has the predicate *flink*til*, taking the prepositional complement as its OBL-TH. The value of OBL-TH is identical with the value of FOCUS-INT, which is indicated by the shared index 8, resulting from the general analysis of *wh*-questions in the grammar.

4 Representing flexibility

Flexible MWEs must be recognizable across different types of syntactic modifications which separate their parts from each other in the sentence. Such modi-

fications include the simple occurrence of other words between the MWE parts, long-distance dependencies like topicalization, relativization and *wh*-question formation, presentative constructions, and various types of passive constructions. When flexible MWEs are treated by means of LFG templates, such modifications are automatically taken care of within the regular grammar. Having both a c-structure and an f-structure representation allows us to capture both the close semantic and functional association between the selecting and the selected words (in the f-structure) and their syntactic independence as different constituents (in the c-structure). We will present the analyses of some cases.

4.1 Intervening words

The simplest case of syntactic modification of an MWE is when other words, typically adverbs, occur between the MWE components. In a verb-second language like Norwegian the sentence subject also frequently breaks up a verb phrase MWE. The treebank example in (40) illustrates the recognition of the predicate *trekke*seg*tilbake* ('withdraw', literally: 'draw oneself back') across several intervening words. The analysis is shown in Figures 13 and 14.

- (40) Da **trekker** jeg **meg** bare stille og rolig **tilbake**.
 then draw I me only silently and calmly back
 'Then I simply withdraw silently and calmly.'

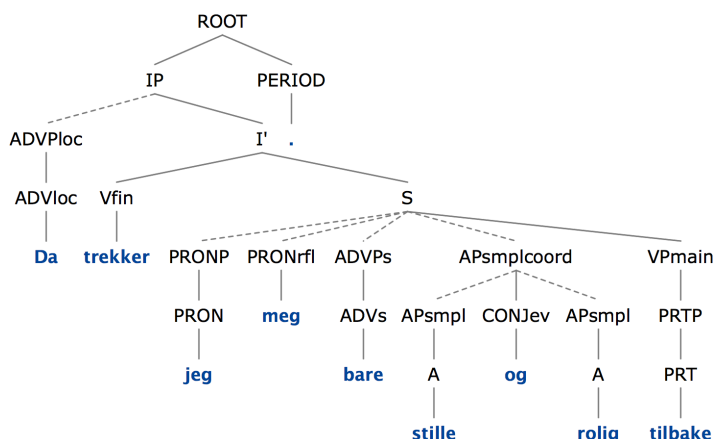


Figure 13: C-structure of sentence (40)

3 Multiword expressions in an LFG grammar for Norwegian

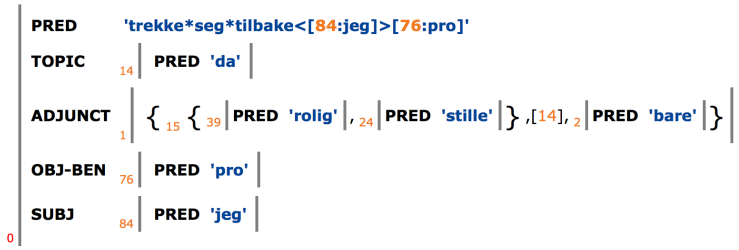


Figure 14: Simplified f-structure of sentence (40)

The mechanism for achieving this lies in the projection architecture of LFG, in which different constituents in c-structure may project the same f-structure, within which dependencies may be formulated. To illustrate we may consider the relevant fragments of the c-structure rules for I', S and VPmain in (41–43).

$$(41) \quad I' \rightarrow \text{Vfin: } \uparrow = \downarrow \\ \quad \quad \quad (S: \quad \uparrow = \downarrow)$$

$$(42) \quad S \rightarrow \text{(PRONP: } (\uparrow \text{ SUBJ}) = \downarrow \\ \quad \quad \quad \quad \quad \quad \quad @\text{SUBJCASE}) \\ \quad \quad \quad \quad \quad \quad \quad [\dots] \\ \quad \quad \quad \quad \quad \quad \quad \text{(PRONrfl: } \{ (\uparrow \text{ OBJ-BEN}) = \downarrow \\ \quad \quad \quad \quad \quad \quad \quad \quad | (\uparrow \text{ OBJ}) = \downarrow \}) \\ \quad \quad \quad \quad \quad \quad \quad [\dots] \\ \quad \quad \quad \quad \quad \quad \quad \text{(ADVPs+: } \downarrow \in (\uparrow \text{ ADJUNCT})) \\ \quad \quad \quad \quad \quad \quad \quad [\dots] \\ \quad \quad \quad \quad \quad \quad \quad \text{(APsmpl: } \downarrow \in (\uparrow \text{ ADJUNCT})) \\ \quad \quad \quad \quad \quad \quad \quad [\dots] \\ \quad \quad \quad \quad \quad \quad \quad \text{(VPmain: } \uparrow = \downarrow) \\ \quad \quad \quad \quad \quad \quad \quad [\dots])$$

$$(43) \quad \text{VPmain} \rightarrow [\dots] \\ \quad \quad \quad \quad \quad \quad \quad \text{(PRTP: } \uparrow = \downarrow \\ \quad \quad \quad \quad \quad \quad \quad \quad (\uparrow \text{ CHECK PRT-VERB}) = c +) \\ \quad \quad \quad \quad \quad \quad \quad [\dots]$$

As explained in Section 2, the equation $\uparrow = \downarrow$ annotated to a rule daughter means that the rule daughter and its mother will project the same f-structure. Thus it can be seen that the *Vfin* daughter of *V'* (the verb) and the *PRTP* daughter of *VPmain* (the particle) will project the same f-structure.

A particle verb presupposes the presence of the required particle in the sentence, and a particle presupposes the presence of a particle verb. This mutual dependency is captured through two features, one feature *PRT-VERB* = +, carried by the verb and required by the rule introducing the particle, and, conversely, one feature *PRT-FORM*, carried by the particle and required by the verb to have the appropriate value. Thus, the constraint equation annotated to *PRTP*, (\uparrow CHECK *PRT-VERB*) = c +, demanding that its f-structure should have a feature *PRT-VERB* = + (i.e., that the verb should be a particle verb), will be satisfied if the finite verb has contributed such a feature to this common f-structure. A similar constraint equation associated with the verb, ((\uparrow *PRT-FORM*) = c prt in (45) below), ensures that the particle has the form required by the verb.

The lexical entry for *trekke* is associated with the relevant frame through an invocation of the template for reflexive verb-particle constructions. Part of the lexical entry for *trekke* is shown in (44). The template has the form shown in (45).

(44) *trekke* V { [...]
 | @(V-SUBJ-OBJrefl-PRT *trekke* tilbake)
 | [...] }

(45) V-SUBJ-OBJrefl-PRT (P prt) =
 @(CONCAT P '* seg '* prt %FN)
 { (\uparrow PRED) = '%FN<(\uparrow SUBJ)>(\uparrow OBJ-BEN)'
 | (\uparrow PRED) = '%FN< \uparrow OBJ)>(\uparrow OBJ-BEN)(\uparrow SUBJ)'
 (\uparrow PRESENTATIVE) = +
 (\uparrow SUBJ PRON-TYPE) = c expl
 ~(\uparrow OBJ DEF) = + }
 @(REFLEXIVE OBJ-BEN)
 (\uparrow CHECK *PRT-VRB*) = +
 (\uparrow *PRT-FORM*) = c prt
 ~(\uparrow *PASSIVE*) = +

The template *CONCAT* constructs the predicate name *trekke*seg*tilbake* as the value of *PRED*. The reflexive occurring with reflexive verbs is analyzed as *OBJ-BEN* (indirect object). The reason for this is that there will be a direct object in the alternative presentative construction with an expletive *det* subject, such as

Det trekker seg tilbake store styrker ‘There are big forces withdrawing’, in which *store styrker* occurs as a syntactic object (OBJ); there can only be one OBJ. The presentative construction is described as the second alternative in the disjunction {...|...} in the template. The reflexive, like the expletive subject, is analyzed as a non-argument, which appears from the fact that it is placed outside the argument list enclosed by ⟨...⟩ in the value of PRED. The features PRT-VRB and PRT-FORM are explained in the discussion of the template in (30).

4.2 Long-distance dependencies

Long-distance dependencies involve syntactic dependencies across an arbitrary number of clause boundaries and comprise topicalization by fronting, relative clauses and *wh*-questions. Such dependencies are handled in the f-structure by means of a special type of equations using regular expressions to specify a set of alternative attribute paths into the f-structure. The term for this mechanism is FUNCTIONAL UNCERTAINTY. The rule in (46) shows a simplified version of the functional uncertainty equation handling the dependency between the topic and some embedded gap further down in the structure.

$$(46) \quad \text{IP} \rightarrow \text{XP}: \quad (\uparrow \text{TOPIC}) = \downarrow \\
\quad \quad \quad (\uparrow \{\text{COMP} \mid \text{XCOMP}\}^* \{\text{SUBJ} \mid \text{OBJ} \mid \text{OBL-TH}\}) = \downarrow \\
\quad \quad \quad [\dots]$$

The first equation annotated to XP in (46) specifies that the f-structure of the XP daughter (\downarrow) is the value of the attribute TOPIC of the f-structure of the IP mother (\uparrow). The second equation specifies that the daughter f-structure is also the value of one of a set of alternative attribute paths. COMP and XCOMP are the attributes of embedded finite and non-finite clauses. The regular expression {COMP | XCOMP}* describes all possible strings over the elements COMP and XCOMP (with repetitions), and the final disjunction specifies the last attribute of the string, enabling the TOPIC to be identical with an embedded SUBJ, OBJ or OBL-TH. We may illustrate with the treebank example in (47) of a prepositional verb *fortelle om* ‘tell about’ in a sentence where the OBL-TH, i.e., the selected prepositional phrase, has been topicalized.

- (47) **Om** dette skal jeg fortelle nå.
 about this shall I tell now
 ‘This I will now tell about.’

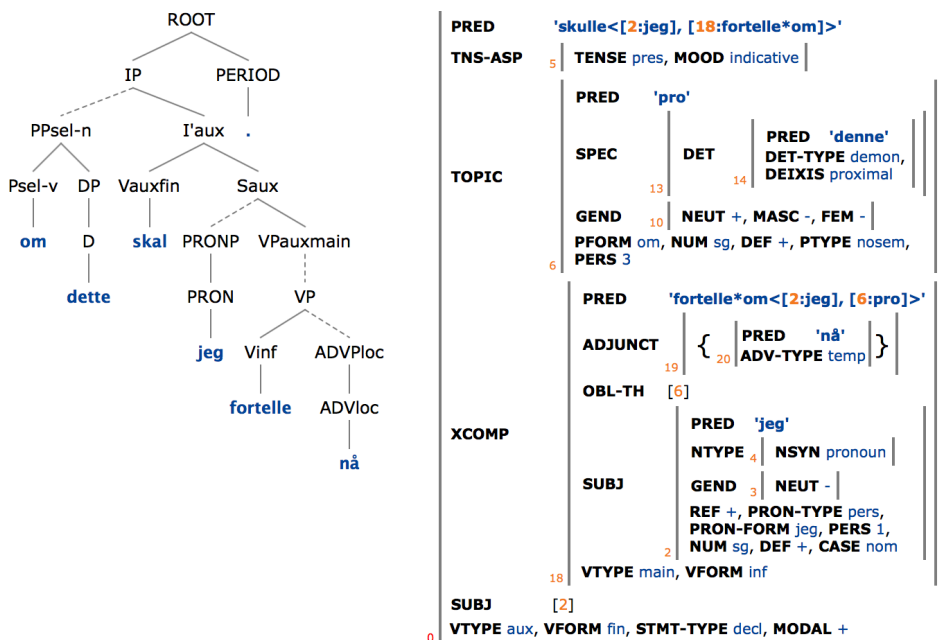


Figure 15: C- and f-structure for example (47)

The analysis of (47) is shown in Figure 15. In the f-structure the value of TOPIC, indexed 6, is also found as the value of OBL-TH in the embedded XCOMP with *fortelle*om* as predicate. Thus the attribute string from the set specified by the functional uncertainty equation in (46) for this example is $(\uparrow \text{XCOMP OBL-TH})$.

4.3 Passive alternations

Passive is another source of verbal MWE modifications, changing the syntactic functions of selected constituents. In LFG passive is analyzed as a lexical phenomenon modifying the value of PRED in a lexical entry for a verb, changing the mapping between argument positions and syntactic functions. In NorGram this is handled by passive templates invoked by the verb templates. The full version of the VP idiom template in (34) for idioms like *holde øye med* ‘keep an eye on’ is shown in (48), where different types of passive alternations are handled.

- (48) VPIDIOM-INDEF OBJ-POBJ (P OP prp) =
 @(CONCAT P '# OP * prp %FN)
 { @(PASS-OBL-TH [(↑ PRED) = '%FN<(↑ SUBJ)<(↑ OBL-TH)>(↑ OBJ)])
 | { (↑ PRED) = '%FN<NULL(↑ OBL-TH)>(↑ SUBJ)<(↑ OBJ)'

| (↑ PRED) = '%FN<(↑ OBL-AG)(↑ OBL-TH)>(↑ SUBJ)(↑ OBJ)' }
 (↑ PASSIVE) = c +
 (↑ PRESENTATIVE-TYPE) = passive
 (↑ SUBJ PRON-TYPE) = c expl }
 (↑ OBL-TH CHECK P-SELFFORM) = prp
 (↑ OBJ PRED FN) = c OP
 ~(↑ OBJ DEF) = +

After the second line there follows a disjunction of two alternatives. The first alternative invokes the template PASS-OBL-TH, taking the predicate-argument structure as a parameter. This template allows the active/passive alternation whereby the OBL-TH, i.e., the complement of the selected preposition (see the discussion of example 25), may be the subject in a passive construction, as in the treebank example in (49).

- (49) De var derimot ikke klar over at de ble holdt øye med.
 they were on the other hand not clear over that they became held eye
 med.
 with
 'On the other hand, they weren't aware that someone was keeping an eye
 on them.'

The second alternative in the main disjunction describes the impersonal (presentative) passive option with an expletive subject, as in the example in (50).

- (50) Det ble holdt øye med dem.
 it became held eye with them
 'Someone was keeping an eye on them.'

The embedded disjunction of two predicate-argument structures in the fourth and fifth lines of the template describes the possibility of including an OBL-AG, i.e., an oblique agent in a prepositional phrase with *av* 'by'. The remaining equations require the passive form of the verb and expletive type of the subject pronoun.

5 Complementation patterns

Verbal MWEs in Norwegian show considerable variation in terms of subcategorizational properties. Like simple verbs, MWEs can have transitivity shifts, take different types of arguments, and take different combinations of arguments. The

verb-particle construction *si opp*, for instance, has both an intransitive reading, as in (51), and a transitive reading, as in (52) and (53). While the shift in transitivity does not significantly affect the semantics of the expression in (52), the shift in (53) leads to a change in meaning.

(51) 150 befal **sier opp**.
150 officers say up
'150 officers resign.'

(52) Hun **sa opp** jobben.
she said up the job
'She resigned from her job.'

(53) Man må **si opp** sjefen for Statkraft.
one must say up the boss for Statkraft
'The head of Statkraft must be fired.'

More precisely, the theme object that is implicit in the intransitive usage in (51) is explicit in (52), while in (53) the object has the semantic role of experiencer instead of theme. The frames V-SUBJ-PRT and V-SUBJ-PRT-OBJ represent the intransitive and the transitive usages of *si opp*. NorGram, being mainly a syntactic framework, has one frame for both transitive readings, leaving semantic roles underspecified.

Most of the verbal MWEs in NorGram are phrasal verbs or VP idioms. Such MWEs have free subjects, so that any argument variation is in the complements.⁷ The lexical entries display a wide range of complementation patterns, one type being MWEs where the verb selects all of its complements. Table 2 presents types of VP idioms in NorGram with only selected complements. In idioms where the verb subcategorizes for only one selected complement, the selected element is either a nominal (O), a prepositional ([P + O]) or a predicative (PC) complement. There is also a type of VP idiom with two selected complements (O + PRT).

Most verbal MWEs in NorGram have free complements in addition to their selected complements. In the VP idiom *legge merke til* 'notice', the verb *legge* 'lay' selects the object *merke* 'mark' in the indefinite form and a prepositional complement which is either nominal, as in (54), clausal, as in (55), or an interrogative clausal complement, as in (56), all headed by the selected preposition *til* 'to'.

⁷The exception to free subjects in VP idioms is expressions with the expletive subject *det* 'it'. However, this type of argument variation is analyzed as a grammatical rather than a lexical selection of the subject and is thus not considered here.

Table 2: Verbal MWEs with only selected complements

Pattern	Example	Lit. translation	Id. translation
V + O	<i>slå følge</i> <i>slå leir</i> <i>ta feil</i> <i>ta fyr</i>	'beat company' 'beat camp' 'take wrong' 'take fire'	'accompany' 'camp' 'be wrong' 'catch fire'
V + [P + O]	<i>gå i oppløsning</i> <i>komme for en dag</i> <i>løfte i flokk</i> <i>legge på svøm</i>	'go in dissolution' 'come for a day' 'lift in flock' 'lay on swim'	'dissolve' 'be revealed' 'join forces' 'start swimming'
V + PC	<i>stå brud</i>	'stand bride'	'get married'
V + O + PRT	<i>sette livet til</i>	'put the life to'	'lose one's life'

- (54) Ingen **legger merke til** mannen som står urørlig og venter.
no one lays mark to the man who stands motionless and waits
'No one notices the man who is standing waiting motionlessly.'
- (55) Ingen **legger merke til** at mannen står urørlig og venter.
no one lays mark to that the man stands motionless and waits
'No one notices that the man is standing waiting motionlessly.'
- (56) Ingen **legger merke til** om mannen står urørlig og venter.
no one lays mark to if the man stands motionless and waits
'No one notices whether the man is standing waiting motionlessly.'

MWEs that subcategorize for different types of complements are represented in the lexicon with one frame for each subcategorization pattern. In the template invocations in (57), P_{OBJ}, P_{COMP} and P_{COMPint} represent the different types of prepositional complements that occur with *legge merke til* in (54), (55), and (56), respectively.

- (57) a. @(VPIDIOM-INDEF_{OBJ}-P_{OBJ} legge merke til)
b. @(VPIDIOM-INDEF_{OBJ}-P_{COMP} legge merke til)
c. @(VPIDIOM-INDEF_{OBJ}-P_{COMPint} legge merke til)

While Table 2 shows different types of selected complements, examples (54–56) illustrate how one MWE may take different types of free complements. In both cases, we see that variation in the complementation is limited for individual MWEs. While *slå følge* ‘accompany’, *gå i oppløsning* ‘dissolve’ and the other examples in Table 2 all have fixed complement structures, *legge merke til* ‘notice’ has three different frames in which only one of the complements varies. To give an impression of the variety of complementation patterns in the lexicon it is thus necessary to turn to the inventory of unique frames, reflected in the number of templates. For instance, NorGram has more than 80 templates for phrasal verbs; these may be grouped into seven main classes according to the types and number of complements (Table 3).

Table 3: Main types of complementation patterns in phrasal verbs in NorGram

Type	Example frame	Example MWE
V + PRT	V-SUBJ-PRT	<i>stryke med</i>
V + PRT + 1 complement	V-SUBJ-PRT-XCOMP	<i>få til</i>
V + PRT + 2 complements	V-SUBJ-PRT-OBJ-OBJ	<i>gjøre etter</i>
V + PPsel	V-SUBJ-POBJ	<i>advare mot</i>
V + PPsel + 1 complement	V-SUBJ-OBJ-PACOMP	<i>erklære for</i>
V + PPsel + 2 complements	V-SUBJ-OBJ-POBJ-PCOMP	<i>vedde med på</i>
V + PRT + PPsel	V-SUBJ-PRT-POBJ	<i>gå med på</i>
V + PRT + PPsel + 1 complement	V-SUBJ-PRT-OBJ-POBJ	<i>venne av med</i>

Table 3 presents the different types of complementation patterns for verb-particle constructions, prepositional verbs, and verb-particle constructions with selected prepositions. The first column in the table is the pattern type, represented in terms of the main complement(s), which may be a particle (PRT), a selected prepositional phrase (PPsel), or both, and the number of additional complements.⁸ Examples of subcategorization frames for each type are given in the second column using template names. The example MWEs, represented in the table with only their fixed components, are instances of the example frames and are discussed in more detail in (58–70).

⁸“Main complement” in this context refers to the selected complement which determines the type of the overall construction, such as PRT in verb-particle constructions.

As Table 3 shows, verb-particle constructions in NorGram may either be intransitive, such as *stryke med* ‘die’ in (58), or have one or two free complements, such as *få noe til* ‘accomplish something’ in (59) and *gjøre noen noe etter* ‘repeat something after someone’ in (60).

- (58) Og vi fortsetter å banke deg til du **stryker med!**
 and we continue to beat you until you stroke with
 ‘And we will continue to beat you until you’re dead!’
- (59) Nå **fikk** han **til** å tenke igjen.
 now got he to to think again
 ‘Now he managed to think again.’
- (60) Ikke mange kunne ha **gjort** ham noe slikt **etter!**
 not many could have done him something like that after
 ‘Not many people could have done what he did!’

The example *få noe til* in (59) is an instantiation of the frame V-SUBJ-PRT-XCOMP, with one free complement in the form of the infinitival complement *å tenke igjen* ‘to think again’. There is one other frame for this particular MWE in the lexicon, with a nominal object instead of the infinitival complement (V-SUBJ-PRT-OBJ). Example (61) illustrates this complement structure.

- (61) Dette er hva du **fikk til**.
 this is what you got to
 ‘This is what you accomplished.’

The lexicon also has a frame for *få til* with two free complements, in the form of an object and an infinitival complement. The difference in the number of complements also yields a difference in meaning, as shown in (62). These should thus be considered different MWEs.

- (62) Hvorfor **får** vi ikke dem **til** å bli?
 why get we not them to to stay
 ‘Why can’t we make them stay?’

The last type of verb-particle construction in Table 3, with two free complements in addition to the particle, is exemplified with the frame V-SUBJ-PRT-OBJ-OBJ. This argument structure, illustrated in (60) for *gjøre noen noe etter*, involves

both an indirect object (OBJ-BEN), *ham* ‘him’, and a direct object (OBJ), *noe slikt* ‘something like that’ (OBJ-BEN is shortened to OBJ in the name of the template). A second frame of this type, which is slightly more complex with a nominal object and a clausal complement (COMP) as well as an expletive subject, is V-SUBJexpl-PRT-OBJ-COMP. The MWE *det faller noen noe inn* ‘something occurs to someone’ in (63) is an example of this frame, literally translating into ‘it falls someone something in’. Except for the expletive subject and the particle, the frame has the same arguments as V-SUBJ-OBJ-COMP for single verbs such as *forklare* ‘explain’. The frame is thus regular in terms of argument structure.

- (63) Det **falt** ikke britene **inn** at særlig mange hadde lyst.
it fell not the Brits in that particularly many had desire
‘It did not occur to the Brits that more than a few should want to.’

In contrast to verb-particle constructions which may be intransitive, prepositional verbs will always have a free complement, introduced by the selected preposition. Prepositional verbs can subcategorize for exactly one prepositional phrase, as in *advare mot noe* ‘warn against something’ in (64), where *mot segregering* ‘against segregation’ is a PPsel.

- (64) Han **advarer mot** segregering.
he warns against segregation
‘He warns against segregation.’

Similar to verb-particle constructions, the prepositional verbs in NorGram can take one or two complements in addition to the selected complement. In (65), *erklære noen for noe* ‘declare someone something’ has one complement, the free object *marken* ‘the mark’, in addition to the selected prepositional phrase *for død* ‘for dead’.

- (65) Der måtte myndighetene **erklære** marken **for** død.
there had to the government declare the mark for dead
‘There the authorities had to declare the (German) mark dead.’

The relevant frame in (65) is V-SUBJ-OBJ-PACOMP, where PACOMP is the selected prepositional phrase. In this case, the preposition *for* takes the adjectival predicative complement *død* ‘dead’. While PPsel is the c-structure category for constituents headed by selected prepositions and may refer to any type of prepositional complement, PACOMP is a syntactic variable that reflects the type of complement.

The final type of prepositional verb in Table 3 is illustrated in (66) with the MWE *vedde noe med noen på noe* ‘bet something with someone on something’.

- (66) Abrams **veddet** en sigarett **med** Browne **på** at det regnet.
 Abrams bet a cigarette with Brown on that it rained
 ‘Abrams bets a cigarette with Brown that it was raining.’

This example is an instance of the frame V-SUBJ-OBJ-POBJ-PCOMP, which has two complements in addition to a PPsel, in this case a free object and a second PPsel. The free object is *en sigarett* ‘a cigarette’. In the first PPsel, which corresponds to POBJ in the subcategorization frame, the selected preposition *med* ‘with’ takes the nominal object *Brown*. In the second PPsel, corresponding to PCOMP, the preposition *på* ‘on’ takes the clausal complement *at det regnet* ‘that it was raining’.

Like prepositional verbs, verb-particle constructions with selected prepositions always subcategorize for at least one free complement. Such constructions can have one complement, as in *gå med på noe* ‘go along with something’ in (67), which is an instance of the frame V-SUBJ-PRT-POBJ.

- (67) I land som Sverige **gikk** fagbevegelsen **med** på de nye tankene.
 in countries like Sweden went the unions with on the new thoughts
 ‘In countries like Sweden the unions went along with the new ideas.’

In (67), the particle is *med* ‘with’, and the free argument *de nye tankene* ‘the new thoughts’ is the complement of the selected preposition *på* ‘on’. The prepositional complement could, however, also be clausal, as in (68), or infinitival, as in (69). With the infinitival complement, there is a shift in meaning from ‘go along with’/‘admit’ to ‘agree’.

- (68) Han vil ikke **gå med** på at hun er utpreget modig.
 he will not go with on that she is exceptionally brave
 ‘He will not admit that she is exceptionally brave.’

- (69) Til Libbys forbauselse hadde Jerry **gått med** på å prøve.
 to Libby’s surprise had Jerry gone with on to try
 ‘To Libby’s surprise, Jerry had agreed to try.’

Verb-particle constructions with selected prepositions may also have two free complements. This is the case for *venne noen av med noe* ‘wean someone off something’ in (70). This example, instantiating the frame V-SUBJ-PRT-OBJ-POBJ, has

the particle *av* ‘off’, the free pronominal object *meg* ‘me’, and the selected prepositional object *med det* ‘with that’. Also here, the prepositional complement may vary. The alternative frame is V-SUBJ-PRT-OBJ-PXCOMP, allowing an infinitival prepositional complement, but in this case yielding no difference in meaning.

- (70) Mor klarte aldri å venne meg av med det.
mother managed never to accustom me off with that
‘Mother never managed to wean me off that habit.’

The examples of complementation patterns for phrasal verbs in NorGram show that the subcategorizational properties of MWEs can be the source of variation both at the syntactic and the semantic levels. We have seen that the main types of complementation patterns in Table 3 are shared by a number of subcategorization frames. Table 4 presents some of the frames that are variants of the type V + PPsel + 1 complement in Table 3 (prepositional verbs with one free complement). The frames are divided into groups of MWEs that share the same or similar types of arguments, resulting in five categories of argument patterning for this type.⁹ While the current section provides only superficial observations about the types of MWE argument patterns in the NorGram lexicon, it seems that a more systematic study of their subcategorizational properties could provide useful information about MWE types and tokens and perhaps also new insights into the relationship between argument patterns and the semantics of MWEs.

6 Conclusion

In this chapter we have shown how the modularization of NorGram makes it possible to integrate MWEs into the LFG analyses in a way that does justice to the proper division of labor between the lexicon and the grammar. On the one hand, each MWE is entered into the lexicon with the information necessary for its idiomatic meaning. On the other hand, the syntactic treatment uses ordinary syntactic rules to the extent that the flexibility of the individual MWE allows.

Up until now MWEs have been severely underrepresented in lexical resources for Norwegian, as they have been for many other languages. The main strategy for NorGram has been to incorporate them into the lexicon and grammar when they are encountered during the construction of NorGramBank. MWEs have thus been added to NorGram in tandem with the development of the treebank. As a natural consequence of the way in which the MWEs are represented

⁹Several frames of this type are not listed here, including frames with expletive subjects and objects and subtypes of clausal complements.

3 Multiword expressions in an LFG grammar for Norwegian

Table 4: Some variants of V + PPsel + 1 complement

Complementation type	Subcategorization frame
Free object and PPsel	V-SUBJ-OBJ-PACOMP
	V-SUBJ-OBJ-POBJACOMP
	V-SUBJ-OBJ-POBJNCOMP
	V-SUBJ-OBJ-POBJ
	V-SUBJ-OBJ-PCOMP
	V-SUBJ-OBJ-PCOMPinf
	V-SUBJ-OBJ-PCOMPint
	V-SUBJ-OBJ-PXCOMP
	V-SUBJ-INDOBJ-POBJ
	V-SUBJexpl-OBJ-POBJ
Reflexive object and PPsel	V-SUBJ-OBJrefl-POBJ
	V-SUBJ-OBJrefl-PCOMP
	V-SUBJ-OBJrefl-PCOMPpat
	V-SUBJ-OBJrefl-PCOMPint
	V-SUBJ-OBJrefl-PXCOMP
PPsel and free nominal complement	V-SUBJ-POBJ-COMP
	V-SUBJ-POBJ-XCOMP
	V-SUBJ-POBJ-OBL
	V-SUBJ-POBJ-OBLBEN
Prepositional reflexive object and free nominal complement	V-SUBJ-POBJrefl-OBJ
	V-SUBJ-POBJrefl-COMP
PPsel and PPsel	V-SUBJ-POBJ-PXCOMP
	V-SUBJ-POBJrefl-POBJ

in the grammar and lexicon, it is possible to search for the various MWE types in the treebank. The wealth of information provided by the LFG representations enables searching for many different properties of the MWEs, and the MWEs may be recovered in all the syntactic variations they occur in. As a result, NorGram-Bank is now an important resource for studying Norwegian MWEs in context.

Acknowledgments

We thank Koenraad De Smedt and two anonymous reviewers for valuable comments and suggestions for improvements. This work was partially financed by the Norwegian Research Council and the University of Bergen through the INESS project.

Abbreviations

=c	symbol in a constraint equation: constrained to be equal to
[P + O]	selected prepositional complement
⟨...⟩	brackets enclosing the list of semantic arguments of a predicate
*	element in the name of the predicate of a lexeme with a selected semantically light element (e.g., a preposition)
#	element in the name of the predicate of a lexeme with a selected semantically heavy element (e.g., a noun, forming an idiom)
%FN	variable over predicate names in a lexical template
→	phrase structure rule expansion
↑	metavariable in an equation, referring to the f-structure of the node immediately dominating the node to which the equation is annotated
↓	metavariable in an equation, referring to the f-structure of the node to which the equation is annotated
~	negation in an equation defining f-structure
INESS	Infrastructure for the exploration of syntax and semantics
LFG	Lexical-Functional Grammar
XLE	the development platform Xerox Linguistic Environment

References

- Baldwin, Timothy & Su Nam Kim. 2010. Multiword expressions. In Nitin Indurkha & Fred J. Damerau (eds.), *Handbook of natural language processing*, Second edition, chap. 12, 267–292. Boca Raton, FL, USA: Chapman & Hall/CRC.

- Bresnan, Joan. 2001. *Lexical-functional syntax*. Malden, MA: Blackwell.
- Butt, Miriam, Helge Dyvik, Tracy Holloway King, Hiroshi Masuichi & Christian Rohrer. 2002. The parallel grammar project. In John Carroll, Nelleke Oostdijk & Richard Sutcliffe (eds.), *Proceedings of the workshop on Grammar Engineering and Evaluation at the 19th international conference on Computational Linguistics (COLING)*, 1–7. Taipei, Taiwan: Association for Computational Linguistics.
- Dalrymple, Mary. 2001. *Lexical functional grammar*. Vol. 34 (Syntax and Semantics). San Diego, CA: Academic Press.
- Dalrymple, Mary, Ronald M. Kaplan & Tracy Holloway King. 2004. Linguistic generalizations over descriptions. In Miriam Butt & Tracy H. King (eds.), *Proceedings of the LFG'04 conference*, 200–208. University of Canterbury, Christchurch, New Zealand: CSLI Publications.
- Dyvik, Helge. 2000. Nødvendige noder i norsk: Grunntrekk i en leksikalsk-funksjonell beskrivelse av norsk syntaks [Necessary nodes in Norwegian: Basic properties of a lexical-functional description of Norwegian syntax]. In Øivin Andersen, Kjersti Fløttum & Torodd Kinn (eds.), *Menneske, språk og felleskap*, 25–45. Oslo: Novus forlag.
- Dyvik, Helge, Paul Meurer, Victoria Rosén, Koenraad De Smedt, Petter Haugereid, Gyri Smørdal Losnegaard, Gunn Inger Lyse & Martha Thunes. 2016. NorGramBank: A ‘deep’ treebank for Norwegian. In Nicoletta Calzolari, Khalid Choukri, Thierry Declerck, Marko Grobelnik, Bente Maegaard, Joseph Mariani, Asunción Moreno, Jan Odičk & Stelios Piperidis (eds.), *Proceedings of the tenth international conference on Language Resources and Evaluation (LREC 2016)*, 3555–3562. Portorož, Slovenia. <http://www.lrec-conf.org/proceedings/lrec2016/summaries/943.html>.
- Kaplan, Ronald M. & Joan Bresnan. 1982. Lexical-Functional Grammar: A formal system for grammatical representation. In Joan Bresnan (ed.), *The mental representation of grammatical relations*, 173–281. Cambridge, MA: The MIT Press.
- Lønning, Jan Tore, Stephan Oepen, Dorothee Beermann, Lars Hellan, John Carroll, Helge Dyvik, Dan Flickinger, Janne Bondi Johannessen, Paul Meurer, Torbjørn Nordgård, Victoria Rosén & Erik Velldal. 2004. LOGON: A Norwegian MT effort. In *Proceedings of the workshop in Recent Advances in Scandinavian Machine Translation*. Uppsala, Sweden.
- Losnegaard, Gyri Smørdal, Gunn Inger Lyse, Martha Thunes, Victoria Rosén, Koenraad De Smedt, Helge Dyvik & Paul Meurer. 2012. What we have learned from Sofie: Extending lexical and grammatical coverage in an LFG parsebank. In Jan Hajič, Koenraad De Smedt, Marko Tadić & António Branco (eds.), *META-RESEARCH workshop on advanced treebanking at LREC'2012*, 69–76. Istanbul, Turkey.

- Maxwell, John & Ronald M. Kaplan. 1993. The interface between phrasal and functional constraints. *Computational Linguistics* 19(4). 571–589.
- Meurer, Paul. 2012. INESS-Search: A search system for LFG (and other) treebanks. In Miriam Butt & Tracy Holloway King (eds.), *Proceedings of the LFG'12 conference* (LFG Online Proceedings), 404–421. Stanford, CA: CSLI Publications.
- Nordgård, Torbjørn. 2000. Nordkompleks: A Norwegian computational lexicon. In *COMLEX 2000 workshop on Computational Lexicography and Multimedia Dictionaries*, 89–92. Patras, Greece: University of Patras.
- Nunberg, Geoffrey, Ivan Sag & Thomas Wasow. 1994. Idioms. *Language* 70(3). 491–538.
- Rosén, Victoria, Koenraad De Smedt, Paul Meurer & Helge Dyvik. 2012. An open infrastructure for advanced treebanking. In Jan Hajič, Koenraad De Smedt, Marko Tadić & António Branco (eds.), *META-RESEARCH workshop on advanced treebanking at LREC'2012*, 22–29. Istanbul, Turkey.
- Rosén, Victoria, Paul Meurer & Koenraad De Smedt. 2007. Designing and implementing discriminants for LFG grammars. In Tracy Holloway King & Miriam Butt (eds.), *The proceedings of the LFG'07 conference*, 397–417. Stanford: CSLI Publications.
- Rosén, Victoria, Martha Thunes, Petter Haugereid, Gyri Smørdal Losnegaard, Helge Dyvik, Paul Meurer, Gunn Inger Lyse & Koenraad De Smedt. 2016. The enrichment of lexical resources through incremental parsebanking. *Language Resources and Evaluation* 50(2). 291–319.
- Sag, Ivan, Timothy Baldwin, Francis Bond, Ann Copestake & Dan Flickinger. 2002. Multiword expressions: A pain in the neck for NLP. In *Proceedings of the 3rd international conference on Computational Linguistics and Intelligent Text Processing* (Lecture Notes in Computer Science 2276), 1–15. Springer.