For a unified treatment of particle verbs

Lionel Clément, Sekou Diao
University of Bordeaux - France

HeadLex 16, July 2016
Polish Academy of Sciences, Warsaw

Introduction

In English, as well as in other languages, there exists a class of verbs composed of a collocation between a verb, and a particle which appears in a distinct syntactic position: the particle verbs.

It is not clear if particles are morphological or syntactic elements (Martin Forst, Tracy Holloway King and Tibor Laczkó (2010)). For the purpose of this paper, we will adopt a syntactic treatment of particles in this paper.

Particle verb constructions can be compositional or idiomatic. In compositional constructions such as 1b, the meaning of the combination of the two morphosyntactic elements is partly predictable from the meaning of each separate element, whereas in idiomatic constructions 1a, the meaning of the combined elements is idiosyncratic, requiring a specific lexical entry for the idiomatic particle verbs.

(1) a. The student **gave** it **up**.
b. The student **moved** the box **up**.

Following the analysis of verbal particles introduced in the English ParGram[1] grammar (Martin Forst, Tracy Holloway King and Tibor Laczkó (2010)), in this paper we try to develop an LFG analysis of particle verb constructions which seems to us to be closer to a general linguistic description of this phenomena and argue in favor of a lexically oriented approach to its formalization.

We use the XLFG (Clément (2016)) parser/framework which provides us with tools to efficiently combine lexical entries as we will show in the next section.

It is well known that compositional particle verb constructions may be productive (Villavicencio (2003)), a fact which can be difficult to handle in an electronic lexicon with a wide coverage.

Our approach seems to gracefully handle this difficulty for computational linguistics: each lexical entry for non-compositional idiomatic particle verb contains only idiomatic information such as predicate argument structure and sub-categorization frame. It allows us to capture the fact that the argument structure of an idiomatic particle verb can differ from the argument structure of the same base verb without a particle.

The major focus of our approach concerns the treatment of compositional particle verbs. To analyze these constructions, we use a ranking mechanism to select the non-compositional analysis by default when such an analysis is present in the lexicon. If such an entry is not found, the parser will generate a compositional analysis of the particle verb construction.

Martin Forst, Tracy Holloway King and Tibor Laczkó note that systematically analyzing particle verbs as idiomatic constructions is a problem for the coverage of computational grammars, as every possible combination of a verb and a particle should explicitly be listed in the lexicon. However, some verb + particle combinations are highly productive and the particle may contribute the same meaning or the same discursive context in each case, it should therefore be more parsimonious if compositional constructions were generated on the fly by the parser.

The approach we have taken to handle particle verb constructions combines these two strategies: we list all idiomatic particle verbs in the lexicon and generate compositional particle verb construction by combining the syntactic information contributed by both the verb and its particle.

[1] Butt et al. (2002)

1 Idiomatic constructions

Idiomatic constructions of particle verbs are those constructions where the meaning and the argument structure of the particle verb can not be derived from composing the meaning and argument structure of the verb and its particle. Idiomatic particle verbs must then be listed in the lexicon.

(2) a. John gave Mary the book.
b. John gave the book to Mary.
c. John gave up playing the piano.
d. John gave up his house.
e. John gave up on her.

As the argument structure of *gave* and *gave up* are different (as illustrated in 2) while the other morphosyntactic information such as tense, aspect, agreement etc.. are shared between the two verbs, only the sub-categorization frame of the particle verb *gave up* is listed in the lexicon. The remainder of the feature-structure of the particle verb is provided by the information in the lexical entry of *gave*. See section 3.1 for a technical account of this approach.

2 Productive constructions

Productive constructions of particle verbs are those constructions where the meaning and argument structure of the particle verb is predictable from composing the meaning and argument structure of the verb and the particle.

These constructions are highly productive in English, especially with adverbial particles such as *up*, *down*, *by* and new uses of verb + particle constructions in a productive setting are regularly appearing in corpora. It is therefore uneconomical to list all the potential uses of productive particle verb constructions in the lexicon.

(3) a. John shot the ball.
b. The pilot shot the plane down.

The approach we have taken to handle the productive case is described in section 3.2.

3 Implementation with XLFG

It[2] is impossible to unify two structures with distinct PRED features. This is the standard way of ensuring that each syntactic function is instantiated no more than once.

This said, it is well known that *complex predicate constructions* is a phenomena where two distinct constituents contribute to the specification of a PRED value.

To model such cases, XLFG supports the concatenation operator "-" that derives a PRED value by the combination of a lexeme and a prefix or a suffix lexeme. As a PRED is given by one lexical entry and corresponds to a specific predication for a verb, a combination between a prefix (resp. a suffix) for the particle and a PRED for the main verb corresponds to a particle verb.

In order to describe only the lexeme of a `PRED` attribute but not the sub-categorization, XLFG allow us to use the attribute `LEXEME` instead of `PRED`. This attribute can combine with a `PRED` to form a *complex predicate* without altering the `PRED` sub-categorization frame. By contrast, the `SUBCAT` attribute is equivalent to the `PRED` attribute minus the lexeme. It allows us to describe only the sub-categorization of a *complex predicate*.

In summary, `PRED`, `LEXEME`, and `SUBCAT` may be combined together to give a complex predicate.

3.1 Idiomatic particle verbs

In the particular case where the meaning of a particle verb is not predictable from its components, namely the main verb and the particle, the sub-categorization may also not be always predictable.

[2]This analysis is based on the 9.8.0 version of XLFG (Clément (2016)).

(4) a. He gave /a concert/a toy to a child/me his phone number/*on painting with oil/*on me/*fishing.
b. They gave up /their personal possessions/*their personal possessions to a child/on painting with oil/on me/fishing.

In such a case, XLFG provides for the possibility to override the sub-categorization given by the `PRED` attribute with the value of the `SUBCAT` attribute. Then, the `SUBCAT` attribute is used to describe the sub-categorization of an idiomatic entry.

In order to introduce a new lexical entry for such a combination, XLFG makes available the `#` symbol followed by the new lexeme as follow:

GIVE_UP [SUBCAT: '< SUBJ, [XCOMP | OBLon] >']

Here, the # symbol allows us to create a lexical entry which has no associated morphological form in the lexicon, but a lexeme which is the result of the combination between the verb's lexeme and the particle lexeme part.

Given the lexical entries for the main verb *to give* and the particle *up* encoded as follow:

give V [PRED: 'GIVE<SUBJ, OBJ, [OBLto | OBL]>', tense: present]
up VERB_PART [LEXEME:' - UP']

The lexeme `GIVE_UP` is generated on the fly by the parser thanks to the concatenation operator "-" presents on the particle's lexeme definition.

The feature-structure resulting from unification is the following:

[PRED: 'GIVE_UP < SUBJ, [XCOMP | OBLon] >', tense: present]

- the lexeme is `GIVE_UP`, the concatenation of `GIVE` and `UP`,
- The sub-categorization is given by the # lexical entry,
- The others features are given by the unification of the three lexical entries (verb, particle, and verbal particle)

Here [3] , in summary, are the various combinations for unification between `PRED`, `LEXEME` and `SUBCAT` attributes in XLFG:

	PRED:'X<Y>Z'	LEXEME:'X'	SUBCAT:'<Y>Z'
PRED:'A<B>C'	PRED:'X⊕A<B U Y>C U Z'	PRED:'X⊕A<B>C'	PRED:'A<Y>Z'
LEXEME: 'A'	PRED:'X⊕A<Y>Z'	LEXEME:'X⊕A'	PRED:'A<Y>Z'
SUBCAT: '<B>C'	PRED:'X<B>C'	PRED:'X<B>C'	none

The lexical entries are ranked[4] in XLFG, and the idiomatic entries marked with a # symbol are preferred. This simple method allows us to override the compositional construction when an idiomatic entry exists.

3.2 Adverbial particle

Another lexical entry is required to encode the adverbial particle. In such a case, the `PRED` value is the predicate of the verb itself without any modification.

fly V [PRED: 'FLY<SUBJ>', tense: present];
up VERB_PART [locative: true];

The result of unification for *fly up* is the following

[PRED: 'FLY< SUBJ >', tense: present, locative: true]

[3] X⊕A is the concatenation between X and A, depending on if X (vs. A) is a prefix or a suffix.

[4] Rank is a value which permits to give precedence to a certain lexical entry over an other, with no statistical basis or formal links to Optimality Theory (we are planning to investigate the use of OT to resolve this issue). For now this value is assigned by the grammar engineer according to the specific language for which he is designing the grammar and relies on his linguistic intuitions.

3.3 An example of XLFG analysis

Given the following sentences to parse:

(5) a. He gave a toy to a child.
b. He gave up on his car.
c. The pilote flew up to 40,000 feet.

A simplified[5] sample of XLFG lexicon entries are the following[6]:

```
flew V [PRED: 'FLY<SUBJ, [OBLto]>', tense: past];
gave V [PRED: 'GIVE<SUBJ, OBJ, [OBLto | OBL]>', tense: past];
up VERB_PART [LEXEME:' - UP'] // A particle which must be combined with another PRED
   | VERB_PART [locative: true]; // A second entry for "up": a locative particle
# GIVE_UP [SUBCAT: '< SUBJ, [XCOMP | OBLon] >']; // A lexical entry for the idiomatic particle verb
```

A simplified example of the XLFG analyses of these sentences are the following:

```
[PRED 'GIVE<SUBJ, OBJ, [OBLto | OBL]>'
 SUBJ [PRED 'PRO']
 OBJ [PRED 'TOY']
 OBLto [PRED 'CHILD']
]

[PRED 'GIVE_UP<SUBJ, [XCOMP | OBLon]>'
 SUBJ [PRED 'PRO']
 OBLon [PRED 'CAR']
]

[PRED 'FLY<SUBJ, [OBLto]>'
 SUBJ [PRED 'PILOTE']
 OBLto [PRED '40,000 FEET']
]
```

4 Conclusion

We have proposed an analysis of particle verbs which allows to construct with the same optimized lexicon non-compositional (idiomatic) and productive compositional particle verb constructions. In the latter case, we do not need any special mechanism to compose a verb and a semantically pertinent particle.

In the case of idiomatic particle verb constructions, a mechanism specific to XLFG was used. This allowed us to propose a preferential choice of the idiomatic expression by adding a special entry in the lexicon.

This way of prioritizing an idiomatic construction, if it exists, and fall back on a compositional construction by default, seems to be coherent with both the FLG formalism and an approach where the lexicon takes a prominent place in the analysis of such phenomena.

The XLFG software allowed us to provide this analysis and to efficiently deal with the technical challenges involved in managing productive constructions of particle verbs in the development of an English LFG grammar.

[5] We do not write the complete feature-structures to make the content more readable.
[6] The XLFG comments start with // symbol.

References

Alsina, Alex, Joan Bresnan, and Peter Sells. 1997. Complex predicates: Structure and theory. In A. Alsina, J. Bresnan, and P. Sells, eds., *Complex Predicates*, pages 1–12. Stanford, CA: CSLI Publications.

Butt, Miriam, Tracy Holloway King, Hiroshi Masuichi, and Christian Rohrer. 2002. The parallel grammar project. In N. J. Carroll and R. Sutcliffe, eds., *Proceedings of the Workshop on Grammar Engineering and Evaluation*, pages 1–7. COLING02.

Butt, Miriam, María-Eugenia Niño, and Frédérique Segond. 1999. *A Grammar Writer's Cookbook*. Stanford, CA: CSLI Publications.

Clément, Lionel. 2016. XLFG Documentation https://hal.archives-ouvertes.fr/hal-01277648. Technical report, LaBRI.

Kaplan, Ronald M., Tracy Holloway King, and John T. Maxwell III. 2002. Adapting existing grammars: The xle approach. In N. J. Carroll and R. Sutcliffe, eds., *Proceedings of the Workshop on Grammar Engineering and Evaluation*, pages 29–35. COLING02.

Martin Forst, Tracy Holloway King and Tibor Laczkó. 2010. Particle Verbs in Computational LFGs: Issues from English, German, and Hungarian . In M. Butt and T. H. King, eds., *The Proceedings of the LFG '10 Conference*, pages 228–248. Ottawa, Canada.

Villavicencio, Aline. 2003. Verb-particle constructions and lexical resources. In *Proceedings of the ACL 2003 Workshop on Multiword Expressions: Analysis, Acquisition and Treatment - Volume 18*, MWE '03, pages 57–64. Stroudsburg, PA, USA: Association for Computational Linguistics.