

# For a unified treatment of particle verbs

Lionel Clément, Sekou Diao

University Bordeaux - France

lionel.clement@labri.fr - sekou.diao@etu.u-bordeaux-montaigne.fr

## The problem

- The student **gave** it **up**.
  - The student **moved** the box **up**.

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)). From now on, 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.

## 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.

- John gave Mary the book.
  - John gave the book to Mary.
  - John gave up playing the piano.
  - John gave up his house.
  - John gave up on her.

In (2), we can see that *gave* takes three arguments

< SUBJ, [OBJ1 | OBJ2], [OBJ2 | OBLto] >

while *gave up* takes two arguments

< SUBJ, [OBJ1 | XCOMP | OBLon] >

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*.

## 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.

- John shot the ball.
  - The pilot shot the plane down.

## The known solutions

A satisfactory solution is the analysis of verbal particles introduced in the English and German ParGram ([2] grammars (Martin Forst, Tracy Holloway King and Tibor Laczkó (2010)).

They propose

- The compositional particule verbs are composed in the syntax.
- Idiomatic constructions are listed in the lexicon and have PRED values which are composed of the particle and the verb.
- The argument structure of these composed constructions may differ from the main verb.
- C-structure rules take into account the morphological analysis of particle verbs for German and Hungarian.

They 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.

## Our solution

Our solution is very close to the one proposed by Forth et al.

The main significant differences were driven by the composition of PRED for productive particle verbs:

The composed PRED is systematic:

- The lemma is the concatenation of the verb stem and particle (similar to ParGram)
- The Sub-categorization is a combination of particle's subcat and verb's subcat.
- The argument structure is a combination of particle's one and verb's one.
- The particle's F-Structure and verb's F-Structure are unified in order to combine the semantic and syntactic properties of each.

We use the XLFG ([4]) 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 ([7]), 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 verb without a particle.

- The student gave it up.
  - The student moved the box up.

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.

## An example of XLFG analysis

A simplified extract of the XLFG lexicon is the following:

flew verb [PRED 'TO\_FLY < SUBJ, [OBLto] >', tense : past]

gave verb [PRED 'TO.GIVE < SUBJ, OBJ, [OBLto|OBL] >', tense : past]

// A particle that must be combined with another PRED

up part [LEXEME ' -\_UP']

up part [locative : true] // A second entry for "up": a locative particle

// A lexical entry for the idiomatic particle verb

# TO.GIVE.UP [SUBCAT ' < SUBJ, [XCOMP|OBLon] >']

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

He gave a toy to a child.

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

He gave up on his car.

|   |       |                                      |   |
|---|-------|--------------------------------------|---|
| [ | PRED  | 'TO.GIVE.UP < SUBJ, [XCOMP OBLon] >' | ] |
|   | SUBJ  | [PRED'PRO']                          |   |
|   | OBLon | [PRED'CAR']                          |   |

The pilot flew up to 40,000 feet.

|   |       |                         |   |
|---|-------|-------------------------|---|
| [ | PRED  | 'TO.FLY<SUBJ, [OBLto]>' | ] |
|   | SUBJ  | [PRED 'PILOTE']         |   |
|   | OBLto | [PRED '40,000 FEET']    |   |

## PRED combination

Here, 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 U A<B U Y>C U Z' | PRED: 'X U A<B>C' | PRED: 'A<Y>Z'   |
| LEXEME: 'A'     | PRED: 'X U A<Y>Z'         | LEXEME: 'X U A'   | PRED: 'A<Y>Z'   |
| SUBCAT: ' <B>C' | PRED: 'X<B>C'             | PRED: 'X<B>C'     | none            |

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

# TO.GIVE.UP [SUBCAT ' < SUBJ, [XCOMP|OBLon] >']

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

give V [PRED 'TO.GIVE < SUBJ, OBJ, [OBLto|OBL] >', tense : present]  
up VERB.PART [LEXEME ' -\_UP']

The feature-structure that has resulted from unification is the following:

- the lexeme is **TO.GIVE.UP**, the concatenation of **TO.GIVE** and **-\_UP**,
- The sub-categorization is given only by the # lexical entry,
- While the other features are given by the unification of the three lexical entries (verb, particle, and verbal particle)

PRED ['TO.GIVE.UP < SUBJ, [XCOMP|OBLon] >', tense : present]

## Adverbial particle

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

fly V [PRED 'TO.FLY < SUBJ >', tense : present]  
up VERB.PART [locative : true]

The combination is the following

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

## References

- Alex Alsina, Joan Bresnan, and Peter Sells. Complex predicates: Structure and theory. In Alex Alsina, Joan Bresnan, and Peter Sells, editors, *Complex Predicates*, pages 1–12. CSLI Publications, Stanford, CA, 1997.
- Miriam Butt, Tracy Holloway King, Hiroshi Masuichi, and Christian Rohrer. The parallel grammar project. In N.Oostijk J. Carroll and R. Sutcliffe, editors, *Proceedings of the Workshop on Grammar Engineering and Evaluation*, pages 1–7, 2002. COLING02.
- Miriam Butt, María-Eugenia Niño, and Frédérique Segond. *A Grammar Writer's Cookbook*. CSLI Publications, Stanford, CA, 1999.
- Lionel Clément. XLFG Documentation <https://hal.archives-ouvertes.fr/hal-01277648>. Technical report, LaBRI, February 2016.
- Ronald M. Kaplan, Tracy Holloway King, and John T. Maxwell III. Adapting existing grammars: The xle approach. In N.Oostijk J. Carroll and R. Sutcliffe, editors, *Proceedings of the Workshop on Grammar Engineering and Evaluation*, pages 29–35, 2002. COLING02.
- Tracy Holloway King Martin Forst and Tibor Laczkó. Particle Verbs in Computational LFGs: Issues from English, German, and Hungarian. In Miriam Butt and Tracy Holloway King, editors, *The Proceedings of the LFG '10 Conference*, pages 228–248, Ottawa, Canada, 2010.
- Aline Villavicencio. 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, 2003. Association for Computational Linguistics.