ISSN 1392-1517. KALBOTYRA. 2005. 55(3)

REPRESENTATION OF ENGLISH PHRASAL VERBS IN ONTOLOGICAL SEMANTICS

Julija Televnaja,

Vilnius University, Institute of Foreign Languages,
Universiteto 5, LT-2000, Vilnius
Tel. 8-681-74840
Email: Julija.Televnaja@vki.vu.lt

Introduction[1]

The goal of the project within the framework of Ontological Semantics was to work out a methodology and to supplement (or, more specifically, to *acquire*) a Natural Language Processing System, Knowledge Base Acquisition Editor, with new data, viz., the class of phrasal verbs. The main challenge to the project was the fact that phrasal verbs are notorious for two characteristic features: their discontinuous order on the syntactic level and non-compositionality on the semantic level. The present article provides an account for the methodological considerations and choices made to fulfill the goals of the project.

Terminological Observations

The recent tendency of synchronic studies seems to be to avoid the term *phrasal verbs*, possibly because of its close association with the requirement for non-compositionality of meaning, and the terms *verb-particle constructions*, or *verb-particle combinations* are preferred instead (Dehé 2002, Gries 2000, Jackendoff 2002, McIntyre 2002, Svenonius 1996; cf., however, with historical studies by, e.g., Denison 1993, Hiltunen 1983, and von Schon 1977). Bolinger pointed out that "being a phrasal verb is a matter of degree" (Bolinger 1971, 6). This statement seems to hold both for earlier semantic classifications (e.g., Bolinger 1971, Makkai 1972) and for more recent studies (e.g., Lindner 1983, O'Dowd 1998). The latter distinguish at least three major classes with respect to meaning (viz., combinations with spatio-temporal, resultative, or idiomatic meaning), but there are numerous instances of complex verbs with gradience in meaning that places them on the boundary between the different classes. Non-compositionality is not seen as a mandatory prerequisite for a verb-particle combination to be included in the analysis, but is rather a classificatory criterion for a relatively small group of complex verbs within a much larger class (Jackendoff 2002, 87, O'Dowd 1998, 67).

---

[1] The results of this study first appeared at *Intelligent Information Systems 2004: New Trends in Intelligent Information Processing and Web Mining (IIPWM 2004). May 17-20, 2004 - Zakopane, Poland.* I sincerely thank Dr. Christine M. Tardy at DePaul University, USA, for her kind help with the English of this paper.

In this discussion, the term *phrasal verbs* (PVs) will be used as it best shows reference to the semantic domain and avoids the question of the status of the particle, which is irrelevant for the present study.

The selection was restricted to the acquisition of the most recurrent combinations based on high-frequency verbs using *Longman Pocket Phrasal Verbs Dictionary* as the initial source, later verifying them with the *FrameNet* selection of PVs. Although the dictionary approach has been widely criticized, it is nevertheless quite acceptable when starting from scratch, which was exactly the present case.

Metalanguage of Ontological Semantics

Ontological Semantics (OS) may be defined as a language-independent computer-oriented theory of meaning. With its first application in the early 80s (the Translator Project), the theory of OS has been further developed, verified, and modified in the numerous subsequent natural language processing (NLP) applications up till the currently run CREST project. The applications of the projects undertaken include, but are not restricted to, machine translation, question answering, information extraction, and text generation.

OS performs the role of a link, or mediator, between natural language and its computer representation within the ontology. With its carefully constructed language-independent inventory of concepts, OS establishes a certain language-universal knowledge of the world (which is constant, or *static*, in the theory) and adjusts the variables, i.e. language-specific *dynamic* processing resources which comprise ecology (i.e., punctuation, word boundary, and spelling formalisms), morphology, and syntax. Each language subsystem (in this case, the class of PVs) whose input data have been given an ontological description is referred to as a *microtheory*, and any application of OS may be considered a microtheory of its own.

Knowledge Base Acquisition Editor (KBAE), the practical implementation of OS, is made of four static resources whose goals are similar to those of lexical semantics. The static resources comprise *ontology*, the main inventory of hierarchically organized language-independent concepts; *Fact Database*, the storage for all remembered instances; *onomasticon*, an inventory of proper names; and *lexicon*, a repository of language dependent lexical entries. In this study, the main focus will fall on the ontology itself.

The ontology is a strictly hierarchical set with the superordinate concept ALL branching into three classes depending on the semantics of a concept: EVENTS, OBJECTS, and PROPERTIES. Each of these classes has its own format of representation, based on the meaning properties of a word rather than its grammatical class. Relations among ontological concepts can be divided into two kinds, hierarchy maintenance and meaning description, and are represented in ontological *slots*. The hierarchical transition form the higher level to the lower one and vice versa is marked by the ontological slots HAS-PARTS and IS-A respectively. The description of meaning for the human consumer is stored in the slot DEFINITION. The computationally relevant information about each concept is distributed among the *facets*, such as the *definition* or *case roles*, which, depending on the kind, are further characterized by *values* or *fillers*. The commonly employed facets in the semantic description of the concept are DEFAULT, SEM, which bear constraint information on the concept, and RELAXABLE-TO. The latter is responsible for constraint relaxation and defines the maximum admissible degree of violation of the concept in SEM. In semantic representation the RELAXABLE-TO facet follows SEM, which is extremely convenient acquiring new entries.

The Fact Database (Fact DB) stores "remembered instances of events and objects" and is sometimes referred to as the "episodic memory" of the ontology (Nirenburg and Raskin 2004, 29). While in the ontology itself the principle of economy is a major criterion, the larger the Fact DB, the better for the ontological system. The large scope of systematically represented and indexed entries in the Fact DB is beneficial in that it shows the whole spectrum of possible instances of one and the same concept and thus helps decrease the number of ambiguous cases. The data of the Fact DB are recorded from the *lexicon.*

The *lexicon* is similar to the conventional dictionary in that it stores all superentries, i.e., sets of all entries sharing the same spelling irrespective of their formal or semantic differences. The lexicon is expanded by manually entering or, using a more technical term, *acquiring* new data. The entry itself consists of several *zones* which store various kinds of information. The zones (parentheses show their abbreviations as given in the templates) used for PVs are these: the Annotations (anno) zone providing the definition (defi), an example, and lexicographer's comments on use peculiarities of the relevant PV. The syntactic structure (syn-struc) zone accounts for the composition of the PV by giving the grammatical category of each member and listing the possible orders by (1) and (2) respectively. Besides, the syn-struc introduces the variables (var) $varX. These are given for coindexing purposes of the syntactic information with the lexical-semantic specification of the semantic structure (sem-struc). The sem-struc stores selectional restrictions which have two functions: on the one hand, they impose constraints on the variables; and on the other, they point to the admissible range of meaning violation.

Below is an example of the PV *make up* as stored in the lexicon:

(1)

```
(make-up-v1
 (anno
  (def "to think of a lie, excuse, or story that is not true in order to deceive someone")
  (ex "He accused the press of harassing his sister and making up stories about her.")
  (comments "pronoun issue; can only go between verb and object"))
 (syn-struc
  (1
  (np ((root $var1) (cat np)))
  (v (root $var0) (cat v))
  (np ((root $var2) (cat np)))
  (prep ((root up) (cat prep))))
  (2
  (np ((root $var1) (cat np)))
  (v (root $var0) (cat v))
  (prep ((root up) (cat prep)))
  (np ((root $var2) (cat np)))))
 (sem-struc
  ((1 2)
  (CREATIVE-COGNITIVE-EVENT
  (agent (value ^$var1) (sem HUMAN))
  (theme (value ^$var2) (sem LIE))
  (modality (type evaluative) (value < 0.9) (scope CREATIVE-COGNITIVE-EVENT)))))
      )
```

Methodological considerations

It has been noticed that combinations with truly non-compositional meaning are rare; however, the large group of (partly) non-compositional combinations has not been sufficiently defined either (Jackendoff 2002). There is a tendency, however, to distinguish between purely prepositional verbs, such as *look at*, *wait for*, and those that are not "purely prepositional", even though their meaning may be "perfectly compositional" (Raskin 2003, p.c.), such as *level off* and *dam up*. It seems necessary to define and divide between the two groups when compiling a dictionary of PVs and to provide a treatment of their possible non-compositionality (an account of the metaphorical meaning in PVs may be found in Makkai 1972, or Nunberg 1994).

In this analysis, it has been suggested to split the entire class of verb-particle combinations into four groups, each with formal or semantic characteristics: purely prepositional verbs, compositional PVs, compositional PVs with extended (metaphorical) meaning, and non-compositional PVs. The first group comprises combinations with literal meaning and is of no interest for the present study. To differentiate between a prepositional verb and a PV, the latter is characterized by at least one of the following criteria:

1) an aspectual feature (constituted by two parameters: *phase* and *iteration*, the former referring to the stage of development of an action and the latter giving a numerical value of the number of times the action takes place) ;
2) a modality feature (revealing attitudinal characteristics and further classified into seven types, each of which may have a value ranging from 0 to 1.0) expressed through an ontological concept;
3) (partially) figurative meaning;
4) non-compositional meaning.

While figurativeness of meaning and non-compositionality are generally accepted in all theories of PVs, the first two criteria constitute characteristic parametric features of OS which allow to approach compositional PVs in a more systematic way. As a result, the class of PVs can be divided into the following subgroups (cf. Baldwin 2002):

1) Compositional PVs:
   a. The particle adds an aspectual feature;
   b. The combination has two admissible orders: *vp p np* and *vp p np* (where *vp p* and *np* stand for *verb phrase*, *particle*, and *noun phrase* respectively);
   c. An additional shade of meaning is manifested in the combination.
2) Metaphorical compositional PVs:
   a. The verb has a metaphorical meaning;
   b. The particle has a metaphorical meaning;
   c. Both the verb and the particle have metaphorical meanings.
3) Non-compositional PVs:
   a. The verb does not preserve its recognizable meaning;
   b. The particle does not preserve its recognizable meaning;
   c. Neither the verb nor the particle preserves its recognizable meaning.

In this approach the concept of the particle is taken broadly, focusing primarily on the type of meaning conveyed by the combination as a whole rather than on formal characteristics of the behaviour of the particle (i.e., separability from the main verb). In this way the polemics as to which order, continuous or discontinuous, is to be considered as basic for separable particles is avoided. This approach is convenient as it prioritizes meaning over form and is familiar in computational approaches (e.g., FrameNet Semantics (Johnson et. al. 2003)). Besides, such treatment alleviates the task of the acquirer and provides for more rapid acquisition of entries, the latter being a fundamental principle of OS (Nirenburg and Raskin 2004). An example of the sem-struc of the compositional PV *pick off* "to shoot and kill people or animals one after the other" (Longman 2001: 257) is given below:

(2)
(KILL
(agent (value ^$var1) (sem HUMAN))
(theme (value ^$var2) (sem ANIMAL))
(aspect (phase begin-continue-end) (iteration multiple))
(modality (type evaluative) (value < 0.5) (scope KILL))))
)

One might question if the acquisition process could be placed on a more formal basis by listing particle meanings separately and thus reducing the bulk of compositional PVs to registering about twenty particle entries. Although the meaning and productivity issues of the particles have been discussed in recent literature (Brugman 1988, Jackendoff 2002, Lindner 1983, McIntyre 2002, Vandeloise 1991 for prepositions), largely descriptive definitions point to the fact that a formal account of particles is still to be developed.

Templates

Initially, four templates to account for the possible syntactic orders of the PVs were proposed. In the acquisition process one of them was found to be misleading (listing an optional *noun phrase* and producing ambiguous representations of PVs) and was rejected. A schematic representation of the admissible orders is given below:

(3)
1) VP NP P / VP P NP
2) VP P NP
3) VP P

Further specifications are possible depending on whether it is the syntactic or semantic structure that needs to be developed. For example, there is a small group of PVs that take a particle and a preposition to convey a more complex meaning, e.g., *look up to*, *make away with*, etc. The preposition in these instances may seem to perform only a grammatical function of a connector; however, a closer analysis reveals that the meaning acquired by the combination used with the preposition is significantly different from the one used without the preposition (Makkai 1972). Even though such cases are on the periphery of the class of PVs, they nevertheless have to be approached with due attention. To account for such complex PVs, it was decided to allow for further modifications within the existing templates (by adding lines where necessary) rather than introduce additional ones. Thus, the syntactic structure of the PV *make away with* above has an extra line with the relevant parameters for the preposition *with*:

(4)
(syn-struc
((np ((root $var1) (cat np)))

(root $var0) (cat v))
(prep ((root away) (cat prep)))
(prep ((root with) (cat prep)))
(np ((root $var2) (cat np))))

Similarly, meaning representations within the semantic structure may be extended by specifying additional parameters. In fact, one of the essential principles in this project was to provide a *theme* facet for every template, which often contains crucial information about the differences between PVs with similar meanings and is particularly true of instances where more than one sense has to be captured within a single template. However, the *theme* facet is a large issue and would need a discussion of its own.

Concluding remarks

The practical goal of this study was to supplement the ontological Knowledge Base Acquisition Editor with a new microtheory of English Phrasal Verbs. The theoretical goal of the work was to find regularities in the structure and semantics of PVs. The main difficulties are encountered at the semantic level and deal mainly with rendering meanings that are neither non-compositional nor compositional, but rather belong to the domain of metaphorically extended meanings and for which OS has not yet developed a theory. Therefore, there are at least two ways for future research.

On the theoretical level, it would be interesting to see whether one of the early assumptions rejected in the course of analysis can suggest a more beneficial treatment of the class of verbs as such. The assumption is to posit *verb + particle* as the unmarked verb structure, which then into *verb + ø* for simplex verbs (traditionally referred to as unmarked) or *verb + particle + particle*, which is a special case of the "heavy" PVs like *make away with*. If this approach shows positive results, the entire verb theory could be presented in a way that is closer to Modern English reality and is not drawn upon the facts of Latin grammar. It would also allow for a more homogenous treatment of the English verb with phrasal verbs standing at the core of the class.

On the more practical, computational level, the approach discussed here has to be tested in real-life task-driven applications to see whether it gives not only satisfactory, but qualitative results. As it is the first experience of entering collocations with non-literal meaning into the ontology, its results will be significant for any future acquisition of multi-word expressions. Besides, the semantic treatment of PVs may assist in the development of the microtheory of metaphor within the framework of Ontological Semantics.

REFERENCES

Baldwin T. 2002. Verb-Particle Extraction. (notes). Retrieved from http://mwe.stanford.edu/20020730.pdf

Bolinger D. 1971. *The Phrasal Verb in English*. Cambridge, MA: Harvard University Press.

Brugman C. 1988. *The Story of Over: Polysemy, Semantics, and the Structure of the Lexicon*. New York: Garland.

Dehé N. 2002. *Particle Verbs in English*. Amsterdam, Philadelphia: John Benjamins Publishing Company.

Denison D. 1993. *English Historical Syntax*. Longman and New York: Longman.

Gries S. 2000. *Multifactorial Analysis in Corpus Linguistics*. New York: Continuum.

Hiltunen R. 1983. *The Decline of the Prefixes and the Beginnings of the English Phrasal Verb*. Turun University, Turku.

Jackendoff R. 2002. English Particle Constructions, the Lexicon, and the Autonomy of Syntax". In Dehé N., Jackendoff R., McIntyre A., Urban S. (Eds.). *Verb-Particle Explorations*. Interface Explorations, 1. Berlin, New York: Mouton de Gruyter, 67-95.

Johnson C.R., Petruck, M.R., Baker C.F., Ellsworth M., Ruppenhofer J., and Fillmore C. 2003. *FrameNet: Theory and Practice*. Retrieved from http://www.icsi.berkeley.edu/~framenet/book/book.html.

Lindner S. 1983. *A Lexico-Semantic Analysis of English Verb Particle Constructions*. Bloomington: Indiana University Linguistics Club.

Makkai A. 1972. *Idiom Structure in English*. The Hague: Mouton.

McIntyre A. 2002. Idiosyncrasy in Particle Verbs. In Dehé, N., Jackendoff, R., McIntyre A., Urban, S. (Eds.). *Verb-Particle Explorations*. Interface Explorations, 1. Berlin, New York: Mouton de Gruyter, 95-119.

Nirenburg S. and Raskin V. 2004. *Ontological Semantics*. Cambridge, MA: MIT Press.

Nunberg G., Sag I.A., and Wasow T. 1994. Idioms. *Language 70*: 491-538.

O'Dowd E.M. 1998. *Prepositions and Particles in English. A Discourse-functional Account*. New York, Oxford: Oxford University Press.

von Schon C. V. 1977. *The Origin of Phrasal Verbs in English*. PhD Dissertation, University of New York at Stony Brook.

Svenonius, P. 1996. *The verb-particle alteration in the Scandinavian languages*. Retrieved March 27, 2004 from http://www.hum.uit.no/a/svenonius/paperspage.html.

Vandeloise C. 1991. *Spatial Prepositions: A Case Study from French*. Chicago: University of Chicago Press.

SOURCES

Knowledge Base Acquisition Editor. 2004.

Longman Pocket Phrasal Verb Dictionary. 2001. *Longman*.

ANGLŲ KALBOS FRAZEOLOGINIŲ VEIKSMAŽODŽIŲ IŠRAIŠKA ONTOLOGINĖJE SEMANTIKOJE

Julija Televnaja

Santrauka

Dažnai pasitaikantys žodinėje ir rašytinėje kalboje anglų kalbos frazeologiniai veiksmažodžiai sukelia rimtų problemų apdorojant natūralią kalbą (Natural Language Processing). Pagrindinės priežastys yra frazeologinių veiksmažodžių nevientisumas, taip pat jų daugiareikšmiškumas. Straipsnyje nagrinėjamas frazeologinių veiksmažodžių išraiškos būdas, remiantis ontologinės semantikos teorija. Aptariami ontologinės semantikos teorijos principai ir jai pritaikytos natūralios kalbos apdorojimo sistemos KBAE (Knowledge Base Acquisition Editor) sudedamosios dalys. Siūloma frazeologinių veiksmažodžių klasifikacija atsižvelgiant į jų reikšmę, taip pat aptariami įmanomi veiksmažodžio ir dalelytės junginių modeliai. Kiekvienas frazeologinis veiksmažodis yra aprašomas vienu iš trijų pagrindinių modelių, kurie atskleidžia sintaksinę ir semantinę struktūras. Sintaksinėje dalyje pažymimi privalomieji junginio elementai bei dalelytės atskiriamumas nuo veiksmažodžio. Semantinėje dalyje pagrindinėmis ontologijos sąvokomis bei papildomais parametrais (linksniais, veikslo bei modalumo kategorijomis) aprašoma leksinė junginio reikšmė. Tikimasi, kad apdorojimo rezultatai gali būti panaudoti metaforos teorijai ontologinėje semantikoje sukurti ir peržiūrėti veiksmažodžio sistemą anglų kalboje.

Įteikta 2005 m. balandžio mėn.