

Verb-particle constructions in a computational grammar of English

Aline Villavicencio

University of Cambridge Computer Laboratory,
William Gates Building, JJ Thomson Avenue,
Cambridge, CB3 0FD, UK
Aline.Villavicencio@cl.cam.ac.uk

Ann Copestake

University of Cambridge Computer Laboratory,
William Gates Building, JJ Thomson Avenue,
Cambridge, CB3 0FD, UK
Ann.Copestake@cl.cam.ac.uk

Abstract

In this paper we investigate the phenomenon of verb-particle constructions, discussing their characteristics and the challenges that they present for a computational grammar. We concentrate our discussion on the treatment adopted in the LinGO ERG. We also analyse how different (conventional and electronic) dictionaries capture them, and the inherent limitations in terms of coverage. Given the constantly growing number of verb-particle combinations, possible ways of dealing with these limitations are investigated, taking into account the regular patterns found in some productive combinations of verbs and particles. One possible way to try to capture these is by means of lexical rules, and we discuss the difficulties encountered when adopting such an approach. We also investigate possible ways of restricting the productivity of lexical rules to deal with subregularities and exceptions to the patterns found.

1 Introduction

In this paper we investigate verb-particle constructions in English and discuss some of the challenges that they pose for a broad-coverage computational grammar. By verb-particle constructions, we mean both idiosyncratic or semi-idiosyncratic combinations, such as *make up*, where the meaning of the combination cannot be straightforwardly inferred from the meaning of the verb and the particle, and also more regular combinations, such as *wander up*. Verb-particle constructions are often highly polysemous: eight senses are listed for *make up* in the Collins Cobuild Dictionary of Phrasal Verbs, for instance. They also show syntactic variation: some particles have a fixed position in relation to the verb, such as *come up*, in *She came up with the idea*, where the particle is expected immediately after the verb, thus the ungrammatical **She came with the idea up*. Others have a more flexible order in relation to the verb, and can equally well occur immediately after the verb, or after another complement.

In terms of usage, verb-particle constructions tend to be thought of as informal: they are sometimes said to be inappropriate in formal writing, and conversely slang is a rich source. Presumably because of this, dialect variation in the use of verb-particle constructions is quite marked: the examples and judgements in this paper are British English, except where otherwise stated.

This paper is organised as follows: in section 2 we analyse the treatment of verb-particle constructions adopted in the Lingo ERG. In section 3 we discuss possible ways of extending this treatment, through the use of lexical rules. After that we analyse how different dictionaries capture them and the coverage they provide. In section 5 we investigate ways of identifying more regular patterns among verb-particle combinations, and in section 6 we discuss the problem of semi-productivity and how the application of these rules needs to be restricted. We finish with some conclusions and future work.

2 Verb-particle constructions in a computational grammar of English

The grammar we will take as our starting point is the LinGO English Resource Grammar (ERG).¹ The LinGO ERG treats verb-particle constructions by means of verb entries which subcategorize for particles. There is a wide range of constructions captured in the grammar, and these vary, for instance, in terms of the subcategorisation frame of the verb-particle combination, the position of the particle and the semantics of the particle. A lexical rule, NP_particle_lr, changes the order of the complements to deal with the NP-particle alternation: its application is controlled by the lexical type of the verb. The selection for the specific particle is via the particle's semantic relation. Particles and prepositions share a lexical entry with an underspecified relation (e.g., **on_rel**), but in the structure for an utterance, the semantic relation for a particle is specialized differently from the independent preposition because of the selection (e.g., to **on_rel_s** as opposed to **on_rel_p**).² For instance, the entry for *wander up* is as follows:

```
wander_up_v1 := v_particle_le &
  [ STEM < "wander" >,
    SYNSEM.LOCAL.KEYS [ KEY _wander_up_rel,
                        --COMPKEY _up_rel_s ] ] .
```

where the semantics of *up* is specialized to the semantically vacuous **up_rel_s**. The scoped logical form for *the dog wandered up* is as follows (ignoring some complications irrelevant for current purposes, such as optional arguments, and an extra event argument for prepositions):

```
prpstn(def(x4,dog(x4),wander_up(e2,x4) ^ up_s(e15,v14)))
```

Note that there is no coindexation between the arguments of *up_s* and *wander_up*. The idea is that selected-for relations, such as *up_s*, are semantically vacuous and can therefore be ignored in the logical form (LF). Contrast this with the logical form for the sentence *The dog wandered along the street*:

```
prpstn(def(x4,dog(x4),def(x12,street(x12),wander(e2,x4) ^ along_p(e2,x12))))
```

An earlier approach in the ERG followed Nerbonne (1995) in actually removing the semantic contribution of the selected-for particle within the process of composition. However, there is now a strong monotonicity assumption underlying semantic composition in the ERG which makes that analysis impossible. An analysis analogous to that of Wechsler (1997) in which the semantic structures for the verb and particle are merged is tempting, but this is also unavailable in the ERG because there is an assumption that the lexical entries contribute individual elementary predications.

There are two main practical problems with the ERG's analysis. The first is that verb-particle entries are never treated as productively formed, which leads to omissions — for instance, while *walk* is in the lexicon, *walk up* is not. This is discussed further below. The second problem concerns semantics. Although the idea that the particle is idiosyncratic and contributes no semantics makes sense for some verb-particle combinations, such as *make up* (in at least some of its uses), it is not so reasonable for the productive cases. For instance, we will argue below that *wander up* can be regarded roughly as:

```
prpstn(def(x4,dog(x4),wander(e2,x4) ^ up_s(e2)))
```

where **up_s** has either a directional or locational/aspectual interpretation, which in both cases can be regarded as qualifying the event of wandering (the semantics is discussed further below). The existing treatment means that the commonality between *wander up* and *walk up* is not captured in the LF, which means that generalizations will be missed in an inference component or in semantic transfer for Machine Translation. Similarly,

¹November 2001 version, available from <http://lingo.stanford.edu/ftp>

²There are some cases in the LinGO ERG where this has not been carried through systematically. The discussion below ignores this, since these seem to be infelicities rather than deliberate distinctions.

there is no semantic connection between *wander* and *wander up*, which also has the disadvantage that it makes it impossible to construct the latter productively.

The semantic vacuity idea also causes some problems for generation, at least when using the chart generator provided in the LKB system (Copestake, 2002). It is unreasonable to assume that a grammar-independent component will be able to produce input LFs with the vacuous selected-for particles, and they thus have to be inserted into an input LF as a separate stage before normal generation with the ERG will work.

3 Regularities in verb-particle constructions: lexical rules

It is often the case that some verb particle combinations form some productive pattern that can be captured, with the particles using a fixed particular meaning to contribute to the meaning of a number of combinations. This is the case of the particle *up*, indicating movement or position, and the verb-particle combinations *jump up*, *get up* and *stand up*. These combinations involve the literal meanings of the verb and particle, and have a transparent semantics.

A simple way of allowing for productive verb-particle combination is to produce an entry similar to the one above from a base verb via a rule that adds particles to the complements list. This is shown schematically below:

$$\left[\begin{array}{l} \mathbf{main_verb} \\ \text{SYNSEM.LOCAL.CAT.VAL.COMPS : } \boxed{} \end{array} \right] \mapsto \left[\begin{array}{l} \mathbf{main_verb} \\ \text{SYNSEM.LOCAL.CAT.VAL.COMPS : } \left[\begin{array}{l} \text{FIRST : } \left[\text{HEAD : } \mathbf{prt} \right] \\ \text{REST : } \boxed{} \end{array} \right] \end{array} \right]$$

This rule simply takes a verb lexeme and adds an extra complement, the particle, to its subcategorization list. The particle contributes a fixed meaning to the meaning of the verb: we discuss the details of the semantics below. This leaves the analysis in the ERG essentially unchanged as far as syntax is concerned. In our current implementation, this rule is restricted to applying only to intransitive and simple transitive verbs, through the typing system, since these are by far the most frequent candidates for a productive approach.

In computational terms, the motivation for capturing productive cases is partly to add coverage, but also to improve reliability of the coding. This rule could be used to generate the verb-particle entry for *wander up* from the entry for *wander*. However, it will of course overgenerate: it needs to be specialized to account for various classes of verb-particle constructions. For instance, even though the particle *up* occurs with a wide range of verbs, it only combines productively with some of these classes. Bame (1999) discusses two such cases: the resultative and the aspectual *up*. For example:

- (1) Kim carried the television up (resultative up)
- (2) Kim ate the sandwich up (aspectual up)

With the resultative *up*, the argument is affected (i.e., at the end of the action the television is *up*). In contrast, the aspectual or completive *up* suggests that the action is taken to some conclusion. Bame's analysis follows Wechsler (1997) in merging semantic structures in order to restrict the verb-particle combinations and also in order to give contrasting semantic structures for these two cases. Unfortunately, as mentioned above, this cannot be directly implemented in the ERG: it also does not lend itself to underspecification, which is important to avoid proliferation of analyses.

One complication, however, is that *up* has a use with some motion verbs in which it simply denotes a contextually salient endpoint to the action:

- (3) Kim was standing in the bottom of the valley. Sandy galloped up.

It is tempting to analyse this as an aspectual *up*, in which the end of the path is indicated. Assuming an approach to event semantics where an activity verb such as *gallop* denotes an event which is underspecified as to whether it includes an end point, the very simple analysis below can be defended:

$$\text{gallop}(e,x) \wedge \text{up-end-pt}(e)$$

where *up-end-pt* is taken as a predicate which is true of terminated events (accomplishments).

An alternative to Bame's account would then be to extend this approach to transitive verbs, where although the *up* also generally has a directional component, the sense of completed path is still present:

$$\text{carry}(e,x,y) \wedge \text{up-end-pt-and-dir}(e) \wedge \text{television}(y)$$

Under this approach, given that the end of the path is *up*, it necessarily follows from the semantic properties of *carry* that the television is also *up*, so it isn't necessary to make the compositional semantics express this directly. We can then utilize a very simple lexical rule, which inherits from the schema given above, but which only takes as input the class of motion verbs with the correct aspectual properties.³ However, we should also note that there is a particle use of *up* which is very similar to the PP argument of a verb such as *put*:

- (4) Kim put the picture up.
- (5) The picture is up.
- (6) Kim put the picture on the table.
- (7) The picture is on the table.

Associating individual particles with subtypes of lexical rules is very similar to the treatment of productive derivational morphology available within the LKB system. For encoding subregularities we use redundancy rules, with the verb-particle lexical entry default inheriting from the result of applying a rule to a verb. This means that it is possible to relate a base verb form with the verb-particle construction derived from it, which means that the latter inherits from the former all the common information, such as inflectional morphology, so that if the base verb is irregular, so is the verb-particle combination. Moreover, the same idea applies to register and dialect information, which is shared between the base verb and the verb-particle combination (e.g. both *piss* and *piss off* are generally perceived as informal and impolite). However, in other respects the treatment of productive verb particle formation is somewhat different, in that it is possible to also group the particles, so that any one verb of a given group could occur with any one particle of a related group. For instance, the movement verbs (*come, go, jump, run, walk,...*) and the location or direction particles (*down, in, out, up, ...*) can be productively combined by a lexical rule that will generate all the possible verb-particle combinations allowed by these groups (*come down, come in, come out, come up, go down, ...*). This is done more stipulatively than in Bame's analysis, in the sense that the types for the classes of verbs and the classes of prepositions are separately defined, but the actual work involved in doing the encoding for the computational lexicon is much the same. We consider how we can acquire these classes in the next sections.

4 Verb-particle combinations in dictionaries

Although it seems intuitively plausible that there is some degree of productive formation of some verb-particle combinations, it is not clear what proportion of verb-particles might be accounted for in this way. We investigated this using several dictionaries and lexicons: the paper versions of the Collins Cobuild Dictionary of Phrasal Verbs (Collins-PV), and of the Cambridge International Dictionary of Phrasal Verbs (CIDE-PV), the electronic versions of the Alvey Natural Language Tools (ANLT) lexicon (Carroll and Grover, 1989)

³The availability of the hierarchy of lexical rules is a strong counter-argument to Ackerman and Webelhuth's (1998) claims that they are unsuitable for capturing this type of phenomenon (see also Ackerman and Webelhuth (1998:162)).

(which was derived from the Longman Dictionary of Contemporary English, LDOCE), the COMLEX lexicon (Macleod et al, 1998), and the Cambridge International Dictionary of English (CIDE+) lexicon. Table 1 shows the number of verb-particle entries for each of these dictionaries.⁴

Table 1: Verb-Particle Entries in Dictionaries

Dictionary	Entries
ANLT	2,649
CIDE-PV	over 4,500
CIDE+	1,433
Collins-PV	over 3,000
Comlex	3,433
LinGO ERG	276

As we can see from these numbers, the coverage of each dictionary varies considerably. There is a common core of verbs that is described in every dictionary. For instance, there are 1,291 verb-particles combinations that are described in CIDE+, ANLT and Comlex, from a total of 13,890 distinct entries obtained from combining these three dictionaries. However, there is also a significant number of entries that is only described in one or other of these dictionaries. There is much less agreement in this respect between dictionaries than for morphologically derived forms, for instance. Given the large number of entries obtained by combining these dictionaries, it is surprising that a considerable proportion (6.5%) of the entries in the LinGO ERG lexicon are not listed in any of these three dictionaries (this proportion would increase if we took subcategorization etc into account).⁵ Most of these are at least semi-compositional, e.g., *crisp up*, *come together*, *tie on*, and were probably omitted from the dictionaries for that reason,⁶ though some others, such as *hack up*, are probably recent coinages.

As the number of verb-particle constructions keeps constantly growing, even though lexicons and dictionaries provide valuable information about these constructions and their characteristics, they are inflexible resources that contain only a subset of verb-particles, and a significant number of combinations is not going to be contained in any of these resources. This is a clear problem, if we want to be able to construct grammars and lexicons that can capture verb-particle constructions in naturally occurring texts, because we cannot rely only on dictionary information.

5 Productivity among verb-particles

It is possible to find regular patterns in some verb particle combinations, where particles use a specific meaning in the combinations. Thus, one possibility for identifying further regularities is to join verbs together into meaningful groups according to the particles they take, so that any one verb of a given group could occur with any one particle of a related group. For each such verb group and associated particle group there would be a lexical rule that would generate the combinations.

To create these groups a first possibility that we explored was to group the verb-particle entries in the ANLT lexicon according to the particles involved. Using the location particles *down*, *in*, *on*, *out* and *up*, which are five of the most common particles according to Collins Cobuild, we expected that they would highlight patterns of verbs that indicate some kind of movement. We identified a group of 17 verbs that occur with all of these particles, resulting in 85 verb-particle combinations out of the 2,649 listed in the ANLT lexicon. It

⁴These figures do not take into account subcategorisation information, where a given verb-particle construction can occur with more than one subcategorisation frame.

⁵The LinGO ERG lexicon was manually constructed with most of the verb-particle entries being empirically motivated by the Verbmobil corpus. It is thus probably reasonably representative of a moderate-size domain-specific lexicon.

⁶The Cobuild Dictionary explicitly states that literal meanings and combinations are not given for all verbs.

is a surprisingly small number of verbs, even taking into account that these dictionaries do not list all literal combinations. Moreover, with three exceptions, all of these verb-particle combinations are listed in the CIDE-PV, but with the great majority (66) not having the desired transparent literal meaning among the possible meanings listed for these verbs in CIDE+. We then investigated the possibility of these verbs belonging to a meaningful group, namely that of movement verbs, since almost all of the verbs seem to imply some form of movement or need a location. To do that, we grouped the verbs according to Levin's (1993) classes. However, apart from 5 that are motion verbs, the remaining 12 verbs do not present any relation and do not fall into a unique class, according to Levin's classes. So, no large patterns could be found in this manner, mainly because these dictionaries do not list many of the verb-particle combinations with transparent meaning.

A second possibility tested was to use Levin's classes themselves as the groups of verbs. To test this idea, we analysed one of Levin's classes, that of *Roll* verbs (class 51.3.1, *bounce, drift, drop, float, glide, move, roll, slide, swing*) and the subset of five direction or location particles (*in, down, on, out, up*). In a manual analysis most of the verb-particles generated were considered acceptable.⁷ However, when trying to find these combinations in the dictionaries we found that even for the most common of these particles, *up*, only 2 of the 9 verb-particle combinations involving these particles were listed in, for instance, the ANLT lexicon.

These results indicate the difficulty of constructing meaningful groups of verbs that present regular patterns. On the whole, our results so far have been somewhat negative, mainly because the dictionaries do not list all the productive combinations, and so we cannot use them either as a means of discovering classes or of filtering entries. A different problem is encountered if we start with a particular classification of verbs. Levin's (1993) verb classes seem to give us, in some cases, a good indication of verb-particle acceptability. For instance, the great majority of pairings of the 9 verbs in the 'roll' class with the common locative particles are acceptable. However, Levin's classes may be too fine-grained and specific to show patterns of verb-particle constructions. On the other hand, a given class that presents a regular pattern may also contain certain elements that provide subregularities and exceptions to this particular pattern, as we discuss in the next section. The use of corpora is more promising, for filtering at least, and an investigation of the automatic extraction of verb-particle combinations from corpora is described by Baldwin and Villavicencio (2002).

6 Restrictions on productivity

Although there are some cases where it appears reasonable to treat verb-particle combination as fully productive (within fairly finely specified classes), there are also cases of semi-productivity. For instance, many verbs denoting cooking processes can occur with aspectual *up*: e.g., *boil up, fry up, brew up, heat up* (although note *cool down* — there is perhaps some directionality involved as well). But some combinations are implausible e.g., *?sauté up, ?microwave up*. In terms of Levin's classification, this cross-cuts the distinction within the class of cooking verbs (Levin 1993:45.3) between those which are also verbs of preparing (26.3) and those which are not, since *fry* and *softboil* are both verbs of preparing, but while *fry up* is acceptable, *?softboil up* is odd. Conversely, neither *microwave* or *stew* are verbs of preparing according to Levin, but *stew up* is acceptable while *?microwave up* is not.

Similar cases of semi-productivity are found in other classes. For instance, while *vomit, spew* and *puke* occur with *up*, *?regurgitate up* seems unacceptable. It is also worth noting that there is a strong constraint against repeating the same particle: so while *throw up* or *chuck up* mean *vomit*, we do not get **throw up up* or **chuck up up*. To take a further example, Bame (1999) gives *Gene banged up the car* as an example of aspectual *up*, but *bang up* does not generally have the relevant meaning in British English (though the example is comprehensible). *smash up* and *bash up* are usual, but *?crash up* and *?damage up* are both at least odd. Some of the constraints that arise may be due to register, others to general blocking principles. The sub-regularities and exceptions within verbal groups might be dealt with by having lexical rules that semi-productively apply to the members of each group, following Briscoe and Copestake (1999). In this way, we have ways of restricting the application of lexical rules in a certain class in order to avoid producing *?softboil up* while allowing *fry up*.

⁷**drop up* is presumably disallowed because of contradictory directional properties. Some pairings with *on* are strange, but it is unclear whether they should actually be blocked.

It is also worth noting that idiomatic uses may have a connection with productive uses of particles. For instance, *cough up* has a productive meaning, but also the idiomatic one ‘to produce (money or information) unwillingly’ (using the definition from CIDE). This example should not involve the same relation as literal *cough*, but arguably at least, the contribution of *up* can be taken as involving the same relation as in *pay up*, *settle up*, *serve up*. From a computational perspective, we want to underspecify meaning rather than proliferate particles in the grammar, but we need to do this in a manner which is compatible with expressing commonalities of meaning for inference or MT.

7 Conclusions

In this paper we analysed possible treatments for verb-particle constructions in a computational grammar of English. The discussion concentrated on the LinGO ERG, and proposed possible extensions to the treatment adopted. We also investigated the status of these constructions in different dictionaries, and how these were used to identify regular patterns among verb-particle constructions, not only to extend coverage but also to improve reliability of the coding. Lexical rules are a possible means of encoding regular patterns, but as there are potential exceptions to the generalisations in these patterns, we also investigated how to restrict the application of these lexical rules.

Further analysis need to be conducted, but the results obtained so far suggest that having a hierarchy of lexical rules to automatically generate verb-particle constructions with transparent meanings, based on groups of verbs and particles presents a reasonable initial solution to the productivity problem. A range of mechanisms is available within the LKB system to allow for different classes of semi-productivity, and although this does not lead to a smooth gradient between productive and non-productive verb-particle combinations, it at least begins to allow for the range of productivity observed by Bolinger (1971) and other authors.

Acknowledgments

This research was supported in part by the NTT/Stanford Research Collaboration, research project on multi-word expressions.

8 References

- Ackerman F. and Webelhuth, G. “*A Theory of Predicates*”. CSLI Lecture Notes Number 76. CSLI Publications, 1998.
- Bame, K. “*Aspectual and resultative verb-particle constructions with up*”. Handout of talk given at Ohio State University Graduate Linguistics Student Colloquium, 1999.
- Baldwin, T. and Villavicencio, A. (to appear) *Extracting the Unextractable: A Case Study on Verb-particles*. In Proceedings of the Sixth Conference on Computational Natural Language Learning (CoNLL 2002).
- Bolinger D. “*The Phrasal Verb in English*”. Cambridge MA: Harvard University Press, 1971.
- Briscoe, T. and Copestake, A. “*Lexical rules in constraint-based grammar*”. Computational Linguistics 25:4, 487-526, 1999.
- Cambridge International Dictionary of English. Cambridge University Press, 1994.
- Cambridge International Dictionary of Phrasal Verbs. Cambridge University Press, 1997.
- Carroll, J. and Grover, C. “*The derivation of a large computational lexicon of English from LDOCE*”. In B. Boguraev and E. Briscoe (eds.) Computational Lexicography for Natural Language Processing, Harlow, UK: Longman. 117-134, 1989.
- Collins Cobuild Dictionary of Phrasal Verbs. Harper Collins Publishers , 1989.

Copestake, Ann. *Implementing Typed Feature Structure Grammars*. CSLI Publications, 2002.

Levin, B. “*English Verb Classes and Alternations - A Preliminary Investigation*”. The University of Chicago Press, 1993.

Macleod, C., Grishman, R. and Meyers, A. “*COMLEX Syntax Reference Manual*. Proteus Project, NYU, 1998.

Nerbonne, John. Computational Semantics—Linguistics and Processing, Shalom Lappin (ed.) *Handbook of Contemporary Semantic Theory*, 461–484, Blackwells, 1995.

Wechsler, S. “*Resultative Predicates and Control*”. In Texas Linguistic Forum 38: The Syntax and Semantics of Predication. Dept. of Linguistics of the University of Texas at Austin, 1997.