18

Verb-particle constructions in a computational grammar of English

ALINE VILLAVICENCIO AND ANN COPESTAKE

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 a wide-coverage HPSG grammar: the LinGO ERG. Given the constantly growing number of verb-particle combinations, possible ways of extending this treatment are investigated, taking into account the regular patterns found in some productive combinations of verbs and particles. We analyse possible ways of identifying regular patterns using different resources. One possible way to try to capture these is by means of lexical rules, and we discuss the dif£culties encountered when adopting such an approach. We also investigate how to restrict the productivity of lexical rules to deal with subregularities and exceptions to the patterns found.

18.1 Verb-Particle constructions in a nutshell

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*, in (1), 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 *tear up*, in (2).

(1) He knew what he wanted and quickly made up his mind.

(2) In a rage she tore up the letter Jack gave her.

The Proceedings of the 9th International Conference on HPSG.
Jong-Bok Kim and Stephen Wechsler (eds.).

Such constructions are often highly polysemous: for instance, eight senses are listed for *make up* in the Collins Cobuild Dictionary of Phrasal Verbs and among them we have:

(3) to form something:
Half the congress is made up of lawyers.

(4) to invent:
He used to make up tales about dragons and fairies.

(5) to prepare something for someone to use it or have it:
They made a bed up for John in the guest room.

They also show syntactic variation, where each combination can take part in several different subcategorisation frames. For example, *add up* can occur as an intransitive verb-particle combination in (6) or as a transitive one in (7).

(6) It's a few calories here and another hundred calories there, and it all quickly adds up.

(7) We need to add these marks up.

Some particles have a £xed position in relation to the verb, such as *come up*, in sentence (8), where the particle is expected immediately after the verb. Thus (9) is ungrammatical.

(8) She came up with the idea.

(9) *She came with the idea up.

Other combinations have a more ¤exible order in relation to the verb, and can equally well occur immediately after the verb, or after another complement, as *eat up* in sentences (10) and (11) exemplify.

(10) John ate up his cereal.

(11) John ate his cereal up.

Besides complements, certain adverbs are also accepted between the verb and the particle, such as *right* in (12).

(12) He came *right* back.

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 of these constructions. 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.

These constructions have been the subject of a considerable amount of interest, and recent work includes Jackendoff (1997), Bame (1999), Gries (2000), Zeller (1999). However, the degree of ¤exibility that they present still poses a challenge, especially when it comes to attempting to capture them in a computational grammar. In this paper we describe some of these challenges and discuss potential solutions. It is organised as follows: in section 18.2 we analyse the treatment of verb-particle constructions adopted in the Lingo ERG, which is used as the basis for the treatment adopted in this investigation. In section 18.3 we discuss possible ways of extending this treatment, through the use of lexical rules. After that we investigate ways of identifying more regular patterns and how different resources can be used to provide this information, so that more speci£c lexical rules can be constructed. In section 18.5 we face the problem of semi-productivity present in the patterns captured by the lexical rules and discuss how the application of these rules can be restricted. We £nish with some conclusions and future work.

18.2 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) (Copestake and Flickinger, 2000) version of November 2001. This version of the LinGO ERG treats verb-particle constructions by means of verb entries which subcategorise for particles. There are 295 entries that belong to 11 types, which de£ne a wide range of verb-particle constructions, 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. In this way, this rule only applies to certain transitive verb-particle combinations, such as *check out*, allowing both "*check* NP *out*" and "*check out* NP" but not to others such as *take around*, only allowing the form "*take* NP *around*". A particular combination is speci£ed in an entry de£ning the relevant aspects of the verb, and where the selection for the speci£c particle is via the particle's semantic relation. 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 attribute SYNSEM.LOCAL.KEYS.KEY speci£es the semantic relation of the combination and SYNSEM.LOCALS.KEYS.–COMPKEY implements the particle selection by specifying the semantic relation of the speci£c particle required.

In the ERG particles and prepositions share a lexical entry with an underspeci£ed relation. For example, *on* is de£ned as follows, where the semantic relation is the general type **on_rel**.

```
on := p_reg_le &
   [ STEM < "on" >,
   SYNSEM.LOCAL.KEYS.KEY _on_rel ].
```

In the structure for an utterance, the semantic relation for a particle is specialized differently from the independent preposition because of the selection de£ned by the relevant verbal entry. Then in the case of *on* as a particle, e.g. in *add on*, the semantic relation is further specialised to **on_rel_s** as speci£ed in the entry for the combination, as opposed to *on* as a preposition which is **on_rel_p**.[1]

In the entry for *wander up* the semantics of *up* is specialized to the semantically vacuous **up_rel_s**, and the scoped logical form for *the dog wandered up* is as follows (there and below we ignore some complications irrelevant for current purposes, such as optional arguments, and an extra event argument for prepositions):

(13) prpstn(def(x4,dog(x4),wander_up(e2,x4) $\wedge$ up_rel_s(e15,v14)))

Note that there is no coindexation between the arguments of **up_rel_s** and **wander_up**. The idea is that selected-for relations, such as **up_rel_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*:

(14) prpstn(def(x4,dog(x4),def(x12,street(x12),wander(e2,x4) $\wedge$ along_rel_p(e2,x12))))

where one of the arguments of **along_rel_p** is coindexed with the event variable of *wander* (e2) and the other with the index of *street* (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 some practical problems with the ERG's analysis. The £rst 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, and the latter could be productively generated from the former. Instead, in the ERG each verb-particle combination needs to be explicitly de£ned in

[1]There 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.

the lexicon, and this is not only labour intensive and time consuming, but given the huge number of existing combinations and of new combinations that are constantly created, there are always going to be those combinations that are not listed in the lexicon. 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 senses), it is not so reasonable for the productive cases. For instance, we will argue below that *wander up* can be regarded roughly as:

(15) prpstn(def(x4,dog(x4),wander(e2,x4) $\wedge$ up_rel_s(e2)))

where **up_rel_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) and could be compositionally added to the meaning of the verb to generate the meaning of the combination. Furthermore, the existing treatment means that the commonality in the directional interpretation between *wander up* and *walk up*, where the semantics of the particle is shared, 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, even though the semantics of verb is shared, there is no semantic connection between *wander* and *wander up*. Ideally we would like to keep recorded the connection between a verb and a related verb particle combination that could be productively derived from it. Moreover, as there is no explicit link between these two forms, it is impossible to construct the latter productively from the former.

The semantic vacuity idea also causes some problems for generation, at least when using the chart generation following Carroll et al. (1999). 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.

18.3 Regularities in verb-particle constructions: lexical rules

In order to extend the treatment for verb-particle constructions in the ERG one possibility is to investigate regularities in these constructions. It is often the case that some verb particle combinations form some productive pattern that can be captured, with the combinations sharing the semantic contribution of the particles. 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 combinations is to produce an entry similar to the one for *wander up* from a base verb via a rule that adds particles to the complements list. This is shown schematically below:

$$\begin{bmatrix} \textbf{main\_verb} \\ \text{SYNSEM.LOCAL.CAT.VAL.COMPS} : \boxed{1} \end{bmatrix} \mapsto \begin{bmatrix} \textbf{main\_verb} \\ \text{SYNSEM.LOCAL.CAT.VAL.COMPS} : \begin{bmatrix} \text{FIRST} : \begin{bmatrix} \text{HEAD} : \textbf{prt} \end{bmatrix} \\ \text{REST} : \boxed{1} \end{bmatrix} \end{bmatrix}$$

This rule simply takes a verb lexeme and adds an extra complement, the particle, to its subcategorisation list. The semantic contribution of the particle is added compositionally to the meaning of the verb to form the semantics of the combination: we discuss the details of the semantics below. For instance, this rule could be used to generate the verb-particle entry for *wander up* from the entry for *wander*. This solution leaves the analysis in the ERG essentially unchanged as far as syntax is concerned.

In computational terms, the motivation for capturing productive cases is partly to add coverage, but also to improve reliability of the coding. This ensures, for instance, that the entries generated are consistently de£ned in terms of the information de£ned for verbs and particles already contained in the lexicon. However, it will of course overgenerate creating ungrammatical combinations. Thus, this rule needs to be specialized to account for various classes of verb-particle constructions that form grammatical combinations. In what follows we discuss some of the classes that form regular patterns.

Even though the particle *up* occurs with a wide range of verbs, it only combines productively with some classes. Bame (1999) discusses two such cases: the resultative and the aspectual *up*. For example:

(16) Kim carried the television up. (resultative up)

(17) 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 intensi£ed and taken to some conclusion (i.e., the sandwich is totally consumed at the end of the action). 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 underspeci£cation, 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:

(18) 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 underspeci£ed as to whether it includes an end point, the very simple analysis below can be defended:

(19) gallop(e,x) $\wedge$ up-end-pt(e)

where **up-end-pt** is taken as a predicate which is true of terminated events (accomplishments), and compositionally added to the semantics of the verb by the presence of the particle.

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:

(20) carry(e,x,y) $\wedge$ up-end-pt-and-dir(e) $\wedge$ 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.[2]

In this case, this rule generates the desired combination and adds to the action described by the verb, the appropriate semantic predicate associated with *up*:

$$\begin{bmatrix} \textbf{motion\_verb} \\ \text{COMPS} : \boxed{1} \\ \text{SEM.RELS.LIST} : \boxed{2} \end{bmatrix} \mapsto$$

$$\begin{bmatrix} \textbf{motion\_prt\_verb} \\ \text{COMPS} : \begin{bmatrix} \text{FIRST} : \begin{bmatrix} \text{HEAD} : \textbf{prt} \\ \text{SEM.KEYS.KEY} : \boxed{3}\ \textbf{up-end-pt-and-dir} \end{bmatrix} \\ \text{REST} : \boxed{1} \end{bmatrix} \\ \text{SEM.RELS.LIST} : \begin{bmatrix} \text{FIRST:}\ \boxed{2} : \begin{bmatrix} \textbf{verb\_pred} \\ \text{ARG0} : \boxed{4} \end{bmatrix} \\ \text{REST} : \begin{bmatrix} \text{PRED} : \boxed{3} \\ \text{ARG0} : \boxed{4} \end{bmatrix} \end{bmatrix} \end{bmatrix}$$

the general relation associated with the entry for *up* is specialised to **up-end-pt-and-dir**, which has both the sense of completed path and the sense of direction. The added particle is also coindexed to the same event variable as the verb, generating the desired semantic effect, schematically:

carry(e,x,y) $\mapsto$ carry(e,x,y) $\wedge$ up-end-pt-and-dir(e)

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*:

(21) Kim put the picture up.

(22) The picture is up.

[2]The 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) page 162).

(23) Kim put the picture on the table.

(24) 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. This allows us to de£ne more £ne-grained details about the combinations such as the particular semantic contribution of a given particle. By capturing regularities using lexical rules, such as the one above, the idea is to obtain a family of lexical rules, organized in a hierarchy. The LKB system also allows the use of redundancy rules to encode subregularities, with the verb-particle lexical entry default inheriting from the result of applying a rule to a verb. Thus, 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 in¤ectional morphology, so that if the base verb is irregular, so is the verb-particle combination (Copestake et al., 2002). For example, from a verb with irregular morphology like *go* it is not only possible to derive *go up*, but also the third person singular present form *goes up* and the past form *went up*, since all the information about *go* can be used when generating these forms. 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 join particles and verbs into groups, so that any one verb of a given verb group could occur with any one particle of a related particle 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 de£ned, but the actual work involved in doing the encoding for the computational lexicon is much the same, with the groups of verbs and the groups of particles belonging to appropriate types and the lexical rules being applied exclusively to the relevant types. We consider how we can acquire these classes in the next sections.

18.4 Productivity among verb-particles

In this section we discuss possible ways of £nding productive patterns. For this task we use two different sources of information: dictionaries and Levin's verb classes (Levin, 1993).

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. One source of information about verb-particles is dictionaries. Moreover, they may also help us uncover some productive patterns in these combinations. For these purposes we investigated the coverage of verb-particle combinations in 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) (which was derived from the Longman Dictionary of Contemporary English, LDOCE), the COMLEX lexicon (Macleod and Grishman, 1998), and the Cambridge International Dictionary of English (CIDE+) lexicon. Table 1 shows the number of phrasal verb entries for each of these dictionaries, including not only verb-particle constructions, but also prepositional verbs.

TABLE 1 Phrasal Verb Entries in Dictionaries

Dictionary	Entries
ANLT	6,439
CIDE-PV	over 4,500
CIDE+	1,433
Collins-PV	over 3,000
Comlex	10,478

As we can see from these numbers, each of these dictionaries has a considerable number of phrasal verb entries potentially providing us with a good starting point for £nding patterns. There are 13,555 phrasal verbs that are described in the ANLT and Comlex lexicons combined, and from this, 3,107 are entries for verb-particle combinations.[3] However, even though there is a common core of verb-particle combinations that is described in every dictionary, the coverage of each dictionary varies considerably. For example, given the large number of entries obtained by combining these dictionaries, it is surprising that a considerable proportion (16%) of the entries in the LinGO ERG lexicon are not listed in any of these two dictionaries (this proportion would increase if we took subcategorization etc into account).[4] Most of these are at least semi-compositional, e.g., *crisp up*, *come together*, *tie on*, and were

[3]These £gures do not take into account subcategorisation information, where a given verb-particle construction can occur with more than one subcategorisation frame.

[4]The 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-speci£c lexicon.

probably omitted from the dictionaries for that reason,[5] though some others, such as *hack up*, are probably recent coinages. Thus, even though there is a signi£cant number of entries that are common among the different dictionaries, it seems to correspond only to a subset of the total number of entries each dictionary has. For instance, from the total number of entries obtained by combining ANLT and Comlex, only 34% of the entries are listed in both dictionaries with the remaining 66% of the total number of entries being exclusive to one or the other of these dictionaries. There is much less agreement in this respect between dictionaries than for morphologically derived forms, for example.

Having this large amount of dictionary data available, we then investigated the possibility of £nding regular patterns in verb particle combinations, more speci£cally those where the particles use a speci£c meaning in the combinations. The idea is that some such patterns could be uncovered by the classi£cation of verbs 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 possible combinations.

In order to create these groups we used the combined information from two of the electronic dictionaries: ANLT and Comlex. We analysed possible combinations listed in these dictionaries involving the location particles *down, in, out* and *up*, which are four of the most common particles according to Collins Cobuild Dictionary of Phrasal Verbs. We identi£ed a group of 42 verbs that occur with all of these particles, resulting in 168 verb-particle combinations out of the 3,107 listed in these dictionaries. Given that these particles are so common and account for 50.20% of the combinations listed in these two dictionaries, this group of 42 verbs is surprisingly small, even taking into account that these dictionaries do not list all literal combinations. However, most of them seem to form valid combinations with the verbs and particles having transparent meanings, and they imply some form of directional movement (e.g. *come, run, bring, drag, send, ...*) or need a location (e.g. *put, lay, ...*).

Even though dictionaries do highlight some tendencies, no large patterns could be found in this manner, mainly because dictionaries tend to list idiosyncratic combinations at the expense of omitting the more productive ones. Since most of the combinations which would be expected to be found with these particles would be more productive ones, the results were somewhat limited. So, we cannot use dictionaries either as a means of discovering productive classes or of £ltering unwanted combinations. Moreover, the number

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

of verb-particle constructions is constantly growing. Thus, if we want to be able to construct wide-coverage grammars that can capture verb-particle constructions in naturally occurring texts, we cannot rely on dictionaries alone.

A second source of information was found in Levin's classes of verbs. We investigated the possibility of using Levin's classes themselves as the groups of verbs. To test this idea, we analysed the combinations generated by some of Levin's classes and the subset of four direction or location particles (*in, down, out, up*). For instance, one of the classes analysed was that of *Roll* verbs (class 51.3.1, *bounce, drift, drop, ¤oat, glide, move, roll, slide, swing*). In a manual analysis of the combinations involving this class most of the verb-particles generated were considered acceptable.[6] These results suggest that Levin's classes are a good starting point for obtaining productive patterns in verb-particle constructions. Moreover, to test the extra coverage obtained over the dictionaries, we investigated how many of these combinations derived from the *Roll* verbs are already listed in the combined ANLT and Comlex lexicons. We found that 64% of these combinations are not listed in the combined lexicons. Even for the most common of these particles, *up*, 6 out of 9 combinations generated from the *Roll* verbs are not listed in the lexicons. These results are encouraging and suggest that were the family of patterns to be implemented it would help us considerably to extend not only the coverage provided by the grammars, but also that provided by the dictionaries.

Levin's verb classes seem to give us, in some cases, a good indication of verb-particle acceptability, with the great majority of pairings of the verbs in the *Roll* class with the common locative particles being acceptable. However, these classes seem to be too £ne-grained and speci£c and it may be the case that the same pattern can be found in several unrelated classes. For example, the verbs of *Manner of Wiping* (class 10.4.1) and those of *Cutting* (class 21.1) seem to follow the same pattern with the aspectual *up*, but there are no links between these two classes in Levin's classi£cation.

These results obtained indicate the dif£culty of constructing meaningful groups of verbs that present regular patterns, using both dictionaries and Levin's classes. The use of corpora to extract verb-particle combinations may contribute to improving these results. An investigation of the automatic extraction of combinations from corpora is described by Baldwin and Villavicencio (2002).

18.5 Restrictions on productivity

Although there are some cases where it appears reasonable to treat verb-particle combination as fully productive (within fairly £nely speci£ed classes), there are also cases of semi-productivity. For instance, many verbs

[6]**drop up* is presumably disallowed because of contradictory directional properties.

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 other combinations are implausible e.g., ?*sauté up*, ?*microwave up*. In terms of Levin's classi£cation, this crosscuts the distinction within the class of *Cooking* verbs (class 45.3) between those which are also verbs of *Preparing* (class 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* sounds odd.

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 frequency with which a given combination occurs may also in¤uence acceptability judgements.

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). The general idea is to have the basic entries in the lexicon augmented with a representation of the rules that can be applied to them. For both the basic and any attested lexical sign that is generated by applying a given rule to the basic entry, a probability could be assigned, based on corpus information. This approach allows the attested forms to be captured, but, while dispreferring the unattested ones, does not prohibit them, since they are assigned very low smoothed probabilities (Briscoe and Copestake, 1999). In this way we can capture regular patterns, while accounting for subregularities and exceptions, with the semi-productive application of lexical rules allowing e.g. *fry up* while avoiding e.g ?*softboil up*.

It is also worth noting that idiomatic uses of the combinations 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 de£nition 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*. By adopting this position, it is possible to use the usual entry for the particle *up*, which is compositionally added to these verbs to produce these idiomatic combinations. This treatment of idioms is along the lines of Riehemann (2001), allowing for commonalities between the non-productive cases. ¿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.

18.6 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. Lexical rules are a possible means of encoding regular patterns, and we investigated the identi£cation of regular patterns among verb-particle constructions using dictionaries and Levin's classes, not only to extend coverage but also to improve reliability of the coding. As there are potential exceptions to the generalisations in these patterns, we also discussed 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 multiword expressions.

References

Ackerman, Farrell and Gert Webelhuth. 1998. *A Theory of Predicates - CSLI Lecture Notes Number 76*. CSLI Publications.

Baldwin, Timothy and Aline Villavicencio. 2002. Extracting the unextractable: A case study on verb-particles. In *Proc. of the 6th Conference on Natural Language Learning (CoNLL-2002)*. Taipei, Taiwan.

Bame, Ken. 1999. Aspectual and resultative verb-particle constructions with Up. Handout for talk presented at the Ohio State University Linguistics Graduate Student Colloquium.

Bolinger, Dwight. 1971. *The Phrasal Verb in English*. Harvard, USA: Harvard University Press.

Briscoe, Ted and Ann Copestake. 1999. Lexical rules in constraint-based grammars. *Computational Linguistics* 25(4):487–526.

Cambridge. 1994. *Cambridge International Dictionary of English*. Cambridge University Press.

Cambridge. 1997. *Cambridge International Dictionary of Phrasal Verbs*. Cambridge University Press.

Carroll, John, Ann Copestake, Dan Flickinger, and Victor Poznanski. 1999. An ef£cient chart generator for (semi-)lexicalist grammars. In *Proc. of the 7th European Workshop on Natural Language Generation (EWNLG'99)*, pages 86–95.

Carroll, John and Claire Grover. 1989. The derivation of a large computational lexicon of English from LDOCE. In B. Boguraev and E. Briscoe, eds., *Computational Lexicography for Natural Language Processing*. Longman.

Collins. 1989. *Collins Cobuild Dictionary of Phrasal Verbs*. Harper Collins Publishers.

Copestake, Ann and Dan Flickinger. 2000. An open-source grammar development environment and broad-coverage English grammar using HPSG. In *Proc. of the 2nd International Conference on Language Resources and Evaluation (LREC 2000)*.

Copestake, Ann, Fabre Lambeau, Aline Villavicencio, Francis Bond, Timothy Baldwin, Ivan Sag, and Dan Flickinger. 2002. Multiword expressions: Linguistic precision and reusability. In *Proc. of the 3rd International Conference on Language Resources and Evaluation (LREC 2002)*. Las Palmas, Canary Islands.

Gries, Stefan T. 2000. *Towards multifactorial analyses of syntactic variation: The case of particle placement*. Ph.D. thesis, University of Hamburg.

Jackendoff, Ray. 1997. Twistin' the night away. *Language* 73:534–59.

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

Macleod, Catherine and Ralph Grishman. 1998. Comlex syntax reference manual, Proteus Project. `http://nlp.cs.nyu.edu/comlex`.

Nerbonne, John. 1995. Computational semantics—linguistics and processing. In S. Lappin, ed., *Handbook of Contemporary Semantic Theory*. Blackwells.

Riehemann, Susanne. 2001. *A Constructional Approach to Idioms and Word Formation*. Ph.D. thesis, Stanford University.

Wechsler, Stephen. 1997. Resultative Predicates and Control. *In Texas Linguistic Forum 38: The Syntax and Semantics of Predication* .

Zeller, Jochen. 1999. *Particle verbs, local domains, and a theory of lexical licensing*. Ph.D. thesis, University of Frankfurt.