Multiword expressions at length and in depth

Extended papers from the MWE 2017 workshop

Edited by
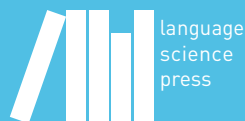
Stella Markantonatou

Carlos Ramisch

Agata Savary

Veronika Vincze

Phraseology and Multiword Expressions 99

language
science
press

Phraseology and Multiword Expressions

Series editors

Agata Savary (University of Tours, Blois, France), Manfred Sailer (Goethe University Frankfurt a. M., Germany), Yannick Parmentier (University of Orléans, France), Victoria Rosén (University of Bergen, Norway), Mike Rosner (University of Malta, Malta).

In this series:

1. Manfred Sailer & Stella Markantonatou (eds.). Multiword expressions: Insights from a multilingual perspective.

Multiword expressions at length and in depth

Extended papers from the MWE 2017 workshop

Edited by

Stella Markantonatou

Carlos Ramisch

Agata Savary

Veronika Vincze

Freie Universität Berlin

Contents

Chapter 1

Identifying senses of particles in verb-particle constructions

Archna Bhatia
Florida Institute for Human and Machine Cognition, Ocala, FL, USA

Choh Man Teng
Florida Institute for Human and Machine Cognition, Pensacola, FL, USA

James F. Allen
University of Rochester, Rochester, NY, USA
Florida Institute for Human and Machine Cognition, Pensacola, FL, USA

In order to attain broad coverage understanding, a system need not only identify multiword expressions such as verb-particle constructions (VPCs), but must compute their meaning. It is not plausible to hand enumerate all possible combinations, although WordNet is an admirable start. This chapter focuses on the identification of senses of particles in VPCs in order to compute the meanings of VPCs – using information obtained from existing lexical resources such as WordNet, and augmenting it with additional knowledge based on linguistic investigation of VPCs identified in terms of generalizations encoded in the TRIPS ontology. The approach consists of first determining compositionality of a VPC based on the information present in WordNet, and then assigning a relevant sense to the particle in a compositional VPC based on the sense classes we have identified and encoded in the TRIPS' computational lexicon. Contributions of the described work are twofold: (1) A discussion of senses of particles in VPCs and corresponding generalizations makes a linguistic contribution. (2) We show how linguistic knowledge can be used to automatically parse sentences containing VPCs and obtain a semantic representation of them. An advantage of the described approach is that VPCs not explicitly found in lexica can be identified and semantically interpreted.

Archna Bhatia, Choh Man Teng & James F. Allen

1 Introduction

To compute deep semantic representations of sentences, we need to pay attention to the richness of lexical meaning. Multiword expressions (MWEs) constitute a significant proportion of the lexicon in any natural language (Moreno-Ortiz et al. 2013). In fact, Jackendoff (1997) estimated the number of MWEs in a speaker's lexicon to be of the same order of magnitude as the number of single words. Thus, it is important to get a good interpretation of MWEs.

This chapter builds on and extends the work reported in Bhatia et al. (2017) with focus on a specific type of MWEs, namely verb-particle constructions (VPCs) VPCs consist of a verb and an adverbial or prepositional particle, e.g., *eat up*, *fade out*, *go on*, *show off*, and *walk down*.[1] Adding every single occurrence of such verb-particle combinations to a lexicon is not efficient nor ideal since knowledge about individual parts (verb and particle) can be leveraged for many of these VPCs as they are interpretable compositionally, e.g., *fly up*.

Other VPCs that indeed are noncompositional require special interpretation, and hence need to be added into the lexicon, e.g., *bring off* 'achieve a goal' and *egg on* 'urge someone for an action that might not be a good idea'. Our work on compositionality of VPCs, described in Bhatia et al. (2017) and developed further here, helps identify VPCs of each type for a proper treatment.

For an inventory of senses for verbs, many lexical resources, such as WordNet (Miller 1995; Fellbaum 1998) and the TRIPS lexicon (Allen & Teng 2017), are available that can be leveraged for interpreting compositional VPC types. In contrast, there is not much for particles except for a few attempts at the semantics of a few particles, such as *up* (Cook & Stevenson 2006) and *out* (Tyler & Evans 2003). However, particles seem to add their own semantics in compositional VPCs and are found to be regular when occurring with verbs in specific verb classes. For example, the particle *up* has a DIRECTION sense when it appears in resultative VPCs with verbs of motion, such as *wander/stroll/go/run up* (Villavicencio 2006). In this chapter, we provide a refined set of senses for particles in VPCs originally presented in Bhatia et al. (2017). We discuss and demonstrate how these senses are identified in compositional VPCs in order to compute meanings of sentences containing VPCs.

For computation of meaning, we use a broad coverage deep semantic parser, TRIPS (Allen et al. 2007), which combines semantic, ontological, and grammati-

---

[1]Note that we focus on the particle usage here, not on the prepositional usage, i.e., a verb followed by a particle not a prepositional phrase. However, there may be an overlap in lexical semantic content (i.e., senses) of the homophonous particles and prepositions, see §4.1.

Draft of October 22, 2018, 16:48

cal information to produce semantic representations.[2] We encode the semantics of particles mentioned above in the TRIPS ontology.[3] The ontology encodes semantic types, the set of word senses and semantic relations that can be used in logical form (LF) graphs. Word senses are defined based on subcategorization patterns and selectional restrictions driven by linguistic considerations. The semantic types in the ontology are, to a large extent, compatible with FrameNet (Johnson & Fillmore 2000). The ontology uses a rich set of semantic features. Unlike WordNet, our ontology does not attempt to capture all possible word senses, but rather focuses on the level of abstraction that affects linguistic processing.

This chapter builds on the work described in Bhatia et al. (2017) in the following ways: Improvements have been made in the classification of compositionality and sense types as well as in the heuristics to automatically identify the types. The evaluation is made more robust with a larger test data set with full coverage of the heuristics. A more thorough analysis has led to the identification of generalizations regarding sense types.

The chapter is organized as follows: Previous work on VPCs is discussed in §2. A classification of VPCs based on their compositionality is discussed in §3 with its feasibility using inter-annotator agreement in §3.1. A set of heuristics to identify different classes of VPCs are presented in §3.2 and their evaluation in §3.3. In §4, we discuss the semantics of particles in VPCs. An inventory of general sense classes for particles used in VPCs is provided in §4.1, which is followed by a brief discussion of manual sense annotations as well as the use of compositionality heuristics for sense identification. In §5, we present various generalizations corresponding to the identified sense classes for the particles, and briefly discuss how a computational lexicon (including a lexicon for particles) is built for the computation of meaning for VPCs. Through examples, we demonstrate the general procedure to compute the meaning of sentences involving compositional VPCs and that the linguistic generalizations are reasonably helpful in the accurate identification of particle senses in VPCs. §6 concludes the chapter.

---

[2]For a more detailed overview of the TRIPS system, refer to Allen & Teng (2017) and Allen et al. (2008).

The TRIPS parser can be accessed at: http://trips.ihmc.us/parser/cgi/parse

[3]The TRIPS ontology can be accessed at: http://www.cs.rochester.edu/research/cisd/projects/trips/lexicon/browse-ont-lex-ajax.html

2 Related work

A lot of computational literature on VPCs focuses on the identification or extraction of VPCs, or on the compositionality of VPCs, as discussed below. There are a few articles dealing with different senses of particles, but they usually focus on only one or two specific particles rather than on a broader coverage of particles. For example, Niinuma (2014) discusses grammaticalization of the particle *away* in English, specifying directional, completive, and continuing or iterative usages of the particle. Ishizaki (2010; 2012) also studies grammaticalization of the particle *away* together with the particle *out*, presenting a classification of VPCs into three categories of fully compositional, partially idiomatic, and fully (or highly) idiomatic.

Vincze et al. (2011) presents the Wiki50 corpus that has 446 VPCs (342 unique types) annotated. Bannard (2002) makes an attempt to identify different types of VPCs in terms of compositionality and builds a (decision tree) classifier to identify the four types. Bannard et al. (2003) also adopt a similar approach for compositionality. As an annotation experiment, they investigate various VPCs to see whether the sense is contributed by the verb and/or the particle. They build four classifiers for automatic semantic analysis of VPCs. Patrick & Fletcher (2004) also have a similar approach, but they focus on automatic classification of different types of compositionality. Unlike our work, in all these works, the focus is on compositionality only, not on the identification of actual senses of the particles.

Cook & Stevenson (2006) discuss various senses for the particle *up* in a cognitive grammar framework, annotate a dataset and perform some classification experiments to identify the senses of *up* in unseen data. As a linguistic study, Jackendoff (2002) provides a very nice discussion of various types of VPCs involving particles such as directional particles, aspectual particles, time-AWAY constructions, and some idiomatic constructions. Our work differs from theirs in having a broader coverage of particles and/or strong emphasis on ontology with respect to the sense classes of the particles and how different particle sense classes relate to verbal ontological classes.

Fraser (1976) mentions semantic properties of verbs affecting patterns of verb-particle combinations, for instance semantically similar verbs *bolt/cement/clam/glue/paste/nail* all can combine with the particle *down* and specify the objects that can be used to join material. Our approach is based on the similar assumption that there are generalizations, such as particles with certain sense classes combine with specific verb classes or ontological classes. Villavicencio (2003) also

adopts the same approach where she tries to encode the information in terms of lexical rules and restrictions, etc. However, her focus is on obtaining productive patterns in VPCs rather than on their interpretation.

Our work also differs from the previous work mentioned above in the following respect: We emphasize the building of complete semantic representations of the sentences, not just the particles' semantics or just the classification of VPCs. Similar to our criteria for compositionality, McCarthy et al. (2003), Baldwin et al. (2003), and Bannard et al. (2003) have looked at distributional similarity as a measure of compositionality of VPCs. In contrast to the approaches focusing on statistical classification based on word/syntax features, our approach (both symbolic and statistical) uses information obtained from existing lexical resources, such as WordNet, for the classification of VPCs. We augment it with additional knowledge based on the linguistic investigation of VPCs identified in terms of generalizations, which we encode into an ontology, in order to compute the semantics of the compositional classes.

3 Classification of VPCs

VPCs have often been classified in terms of their compositionality (i.e., whether all constituents of a VPC, the verb and the particle, contribute their simplex meanings to the overall semantic content of the VPC). The classes fall somewhere between fully compositional VPCs, e.g., *fly up*, and fully idiomatic VPCs, e.g., *egg on*. For example, see Fraser (1976), Chen (1986), O'Dowd (1998), Dehé (2002), and Jackendoff (2002).

We also classify VPCs into two types, compositional VPCs and noncompositional VPCs. The difference between the two classes is that the meaning of a compositional VPC is the sum of the meanings of its parts (the verb and the particle) whereas a noncompositional VPC has semantic content which is not contributed by the individual constituents (i.e., the verb and the particle).

The compositional VPCs can be further classified into three subtypes: symmetrically compositional VPCs, light particle compositional VPCs (LP-compositional VPCs), and light verb compositional VPCs (LV-compositional VPCs), based on the type of semantic content contributed by each of the constituents. Symmetrically compositional VPCs refer to VPCs in which both constituents, the verb and the particle, contribute their simplex meanings (their lexical-semantic content). For example, in *Debris **flew up** and hit the window in the furthest unit*, the senses for the verb *fly* (e.g., in WordNet, sense fly%2:38:00) as well as the particle *up* (e.g., in WordNet, sense up%4:02:00) combine together to provide the mean-

ing of the VPC *fly up*.[4] We distinguish the other two compositional VPC types from the symmetrically compositional VPCs only in the aspect that in the other two types, the particle or the verb have a relatively lighter contribution than the other constituent which adds its regular lexical-semantic content.[5]

LP-compositional VPCs involve particles which, instead of contributing a preposition like lexical-semantic content, contribute aspectual information to the VPCs. Verbs contribute most of the lexical-semantic content in such VPCs. For example, in *Susan **finished up** her paper* (Bannard & Baldwin 2003), the verb *finish* contributes its regular lexical content (e.g., in WordNet, sense finish%2:30:02). However, the particle *up*, instead of contributing its regular lexical-semantic content, adds aspectual information that the action was completed (i.e., the COMPLETELY sense in our sense inventory). See §4.1 for the specific senses of particles.

LV-compositional VPCs involve light verbs which carry bleached meaning compared to regular verbs, e.g., CAUSE and BECOME. The particles contribute their regular lexical-semantic content to the VPCs's semantics. For example, in *The thief **made away** with the cash*, the particle *away* contributes its regular meaning (e.g., WordNet sense away%4:02:00), but the verb *make*, instead of contributing its regular meaning (e.g., WordNet sense make%2:36:01), adds a bleached meaning (e.g., cause to be). For details on the procedure to compute meanings of sentences with compositional VPCs, see §5.

In a noncompositional VPC, the sum of meanings of individual parts (the verb and the particle) may not completely account for the meaning of the VPC or may not account for it at all. Let's consider a few examples. In *They **turned away** hundreds of fans*, the VPC **turned away** is noncompositional despite the fact that the individual constituents' semantic content is reflected in the semantics of the VPC, since the VPC has additional content ('refuse entrance or membership') besides those contributed by the constituents. This additional content is not inferable from the individual parts and needs to be included in the lexical entry for the VPC. Another example of a noncompositional VPC is idiomatic usages. For example, the VPC **egged on** in *John wouldn't have done the dangerous experiment if his brother hadn't **egged** him **on*** involves idiosyncratic content that is not inferable from the parts and hence needs to be encoded in the lexicon. Some idiomatic usages may involve certain generalizations which may aid in interpre-

---

[4]For this study, we have used WordNet version 3.0. The numbers appearing after the symbol % in the WordNet senses represent the sense keys in WordNet.

[5]The term "light particle" is used in analogy with the term "light verb" which is commonly used in the literature for verbs with bleached content.

tation of the VPC. For example, two generalizations regarding the interpretation of the noncompositional VPC **take up** are (i) it takes the sense 'starting to do an activity' when it appears with activities as direct objects, e.g., *She **took up** photography/swimming* [activities] and (ii) it takes the sense 'assume a responsibility' when it appears with positions/responsibilities as direct objects, e.g., *She **took up** her position* [responsibility/position].

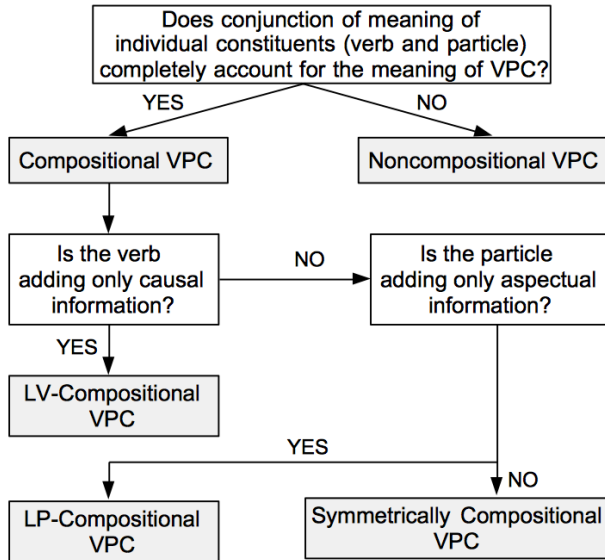For identification of the compositionality type for a given VPC, one may adopt tests as in Figure 1.

Figure 1: Tests to identify compositionality type

3.1 Human agreement on coarse-grained classification of VPCs

From among all the VPCs for which WordNet has an entry, we automatically extracted 50 random VPCs such that four particles, namely *up*, *down*, *out*, and *away*, were represented in the extracted VPCs. Since a VPC may have both compositional and noncompositional usages in different contexts (represented by different word senses or synsets in WordNet), we restricted the assignment of annotation label for a specific VPC to only one label by considering a single synset from WordNet for annotations. Some of the WordNet synsets do not have an ex-

ample with the exact VPCs.[6] We restricted the automatic extraction of VPCs to only those VPCs for which WordNet had a synset which included the VPC in its example. This synset was presented to the annotators together with the VPC to be annotated and the example usage.

These test VPCs were manually annotated by three annotators for compositionality labels. Fleiss' kappa score (Fleiss 1971) was used to test inter-annotator reliability. The three sets of annotations achieved a score of 0.651, an intermediate-good score.[7] We also created the gold annotations from the three sets of manual annotations. In cases of disagreement, the three annotators discussed reasons for their decisions and arrived at a consensus to create the gold annotations for the VPCs. The distribution of compositional vs. noncompositional VPCs in the automatically extracted 50 test VPCs was 60% vs. 40%. In terms of the specific compositionality types, the distribution was 30% for symmetrically compositional cases, 30% for LP-compositional, and 40% for noncompositional cases. Note that the 50 randomly extracted test VPCs did not have an instance of an LV-compositional VPC.

In the manual annotations, there were 37 VPCs out of 50 for which there was full agreement among all three annotators on the coarse-grained labels for compositionality type. Since there were only two coarse-grained labels (compositional and noncompositional), for the rest of the 13 VPCs, there was agreement between two out of three annotators for the compositionality label. The annotators disagreed more on the compositional cases (76.92% of the disagreements were on the compositional VPCs) than on the noncompositional ones (23.08%).

In terms of VPCs with certain particles, VPCs with the particle *down* were found to be the most challenging for the annotators. For 42.86% of the usages of VPCs with *down* (3 out of 7 usages) among the 50 test cases, the annotators did not agree on the annotation label. On the other hand, VPCs with the particle *up* were found to be the least challenging. For only 14.29% of the cases (3 out of 21 usages), the annotators disagreed on the compositionality type. This may be due to the fact that VPCs with the particle *up* have a higher frequency than VPCs with other particles (Villavicencio 2006), such as *down*, and hence users have relatively better intuitions about VPCs with the particle *up* compared to VPCs with the particle *down*.

---

[6]It may have examples which use related verbs or related VPCs instead, or it may not have an example at all. For example, the WordNet entry *turn_away%2:38:02* for the VPC **turn away** does not have any example with the VPC itself.

[7]NLTK's agreement package was used to calculate the Fleiss' kappa score which showed intermediate-good agreement. Cohen's kappa score (Cohen 1960) was also calculated using the same package showing substantial agreement, also with a score of 0.651.

3.2 Heuristics for compositionality of VPCs

As a first step toward an interpretation of VPCs, we need to determine whether a given VPC is compositional or not. To perform this task automatically, we employ a number of heuristics that make use of the rich inventory of hierarchically organized word senses (i.e., synsets) in WordNet which contains over 100,000 words including 64,188 multiwords. Heuristics 1–6 below are used to identify compositional VPCs, whereas heuristic 7 indicates non-compositionality.

1. If the verb is among the list of light verbs, and WordNet does not have an entry for the VPC, it is LV-compositional. For example, the VPC *make away* uses the light verb *make* and the VPC does not have an entry in WordNet.

2. Given a VPC, if heuristic 1 does not apply and WordNet has an entry for the verb as well as for the particle, but no entry for the VPC, VPC is compositional (LP-compositional or symmetrically compositional). For example, *fly* with the sense key fly%2:38:01 as well as *up* with the sense key up%4:02:00 appears in WordNet, but *fly up* does not appear in any synset in WordNet.

3. If WordNet has the VPC as well as the verb in the same synset, VPC is LP-compositional. For example, the VPC *sort out* (sort_out%2:31:00) and the verb *sort* (sort%2:31:00) both appear in the same synset in WordNet.

4. If WordNet has the verb in the VPC as a hypernym for the VPC, VPC is either symmetrically compositional or LP-compositional. For example, compositional VPC *push up* (push_up%2:38:00) has the verb *push* (push%2:38:00) as its direct hypernym.

5. If WordNet has the verb in the definition (in its base or inflected form) of the synset where the VPC appears, the VPC is either symmetrically compositional or LP-compositional. For example, the compositional VPC *move up* (move_up%2:38:00) has the verb *move* in its definition *move upwards*.

6. If WordNet has the relevant VPC as well as another VPC with the verb replaced with another verb in the same synset, the VPC is compositional (either symmetrically compositional or LP-compositional or LV-compositional). For example, the compositional VPCs *pull out* (pull_out%2:35:00) and *rip out* (rip_out%2:35:00) appear in the same WordNet synset.

7. If none of the above heuristics apply, the VPC is noncompositional. For example, none of the above heuristics apply to the idiomatic VPC *catch up* (catch_up%2:38:0).

3.3 An evaluation of the heuristics for compositionality of VPCs

For an evaluation of the heuristics, we used two test sets: (i) Test Set 1: the test set consisting of the same 50 randomly extracted VPCs that were used to calculate inter-annotator agreement scores mentioned in §3.1, and (ii) Test Set 2: a test set consisting of 653 VPCs created using the VPCs on the first page of each of the four English wiktionary entries with the title "Category: English phrasal verbs with particle" for particles *up*, *out*, *down*, and *away* respectively.

A Python implementation of the heuristics was applied to the test VPCs in both test sets to assign them a compositionality label. Heuristics 1–6 identify compositional VPCs, whereas heuristic 7 identifies noncompositional VPCs. Note that together these heuristics have full coverage, i.e., a prediction is made for each of the VPCs in the test sets. Heuristics 1 and 2 apply to VPCs for which WordNet does not have an entry. Heuristics 3-6 apply to VPCs for which WordNet has an entry. Heuristic 7 applies when none of the heuristics 1-6 apply.

Since Test Set 1 consisted of VPCs that were randomly extracted from Word-Net, only heuristics 3–7 could be evaluated on Test Set 1. Heuristics 1 and 2 could not be evaluated using this test set since these two heuristics apply to VPCs that are not included in WordNet. In order to test heuristics 1 and 2, we used Test Set 2 which has VPCs that may or may not be included in WordNet.

Test Set 1 has manually created gold annotations. Hence, the labels assigned by the heuristics were tested against them for the VPCs in Test Set 1. For Test Set 2, on the other hand, we examined the heuristics-assigned labels corresponding to heuristics 1 and 2 manually.

An evaluation of the heuristics on Test Set 1 is presented in Table 1. Heuristics 3, 4, and 5 (used to identify compositional VPCs) performed perfectly in identifying compositionality of VPCs whereas heuristic 6 (also intended to identify compositional VPCs) did not have as high precision. Two out of the three cases where heuristic 6 made an incorrect prediction, however, involved noncompositional VPCs for which the individual constituents' meaning was also reflected in the semantics of the VPCs even though they also had additional content that was not inferable based on the constituents alone. The fact that the constituents' semantics was reflected in the VPCs' semantics was what heuristic 6 had captured. Heuristic 7 (the only heuristic used to identify noncompositional VPCs)

also had a lower but better than chance performance in comparison to the other heuristics.

Table 1: Evaluation of heuristics 3–7 using Test Set 1 (50 test cases)[a]

Heuristic #	Precision for the heuristic (%)	Coverage of the heuristic (%)
3	100	16
4	100	6
5	100	12
6	76.92	26
7	58.62	58
Overall (heuristics 3–7)	70	100

[a]We define precision as Cn/Tn and coverage as Tn/N, where N is the corpus size, Tn is the sample size that heuristic n is applicable to, and Cn is the number of correct assignments it makes.

Overall, heuristics 3–7 had a precision of 70% in assigning a label (compositional or noncompositional) to a VPC. Whenever multiple heuristics applied to a VPC (14% of the test cases), they were always correct in their prediction. This suggests that presence of multiple characteristics of compositionality (represented by the heuristics) may be a reliable indicator of compositionality. In 86% of the cases, a single heuristic identified the VPC as representative of its own category (compositional or noncompositional). Regarding the relatively lower performance of heuristic 7, since noncompositonal VPCs carry additional information not contributed by individual constituents (common sense knowledge or idiosyncratic information), it may not be observable and hence is not captured using a heuristic as easily or directly.

More investigation is required into heuristics 6 and 7 to see how they can be modified to make them better usable for identification of the compositionality type of a VPC. For example, a future step may be to examine if certain generalizations exist in terms of verb types and particle types that can help us further determine when the cases identified by these heuristics are compositional or noncompositional.

As a result of the examination of the heuristics' output, we noticed a few more indicators which could also be incorporated into the heuristics to improve their performance in the future. For example, if the word *completely* or *thoroughly*

appears in the definition of a synset of a VPC, the particle in the VPC may carry the aspectual sense COMPLETELY. Hence, it could be labelled as a compositional VPC (specifically, as an LP-compositional VPC).

As mentioned above, for an evaluation of the heuristics 1 and 2, we used Test Set 2 that consisted of 653 VPCs, a subset of the VPCs mentioned in the English wiktionary. The Python implementation of the heuristics mentioned above was used to assign compositionality labels to the VPCs in Test Set 2. Out of the 653 VPCs, heuristic 1 applied to only 2 VPCs (0.3% of the test items) and heuristic 2 applied to 280 VPCs (42.88% of the test items). Heuristic 7 applied to 9 other VPCs that did not have a WordNet entry. Overall, 44.56% of the test cases did not have an entry in WordNet. Heuristic 2 covered 96.22% of these cases with reasonable precision (80%, evaluated on 15 randomly selected test VPCs to which heuristic 2 applied).

Next, we move on to the semantics of particles in VPCs. We will focus mostly on the compositional VPCs for the rest of this chapter.

4 Semantics of particles in VPCs

As mentioned in §3, particles contribute to the overall semantics of compositional VPCs. In order to study the contribution of particles in VPCs, in our prior work (Bhatia et al. 2017), we conducted an investigation of VPCs consisting of verbs in the ontology class ONT::EVENT-OF-CAUSATION in the TRIPS ontology. This class consisted of 1383 words with verb senses (and a total of 1784 verb senses of those words). Our investigation consisted of combinations of these verbs with the following particles (wherever the combinations appeared as VPCs): *across*, *away*, *by*, *down*, *in*, *into*, *off*, *on*, *out*, *over*, *through*, and *up*.[8] We searched for examples for each of the combinations using Google and manually went through each of the examples to check for a number of properties. For example, we checked if any of the verb or the particle contributed to the overall meaning of the VPC. We identified the senses particles had in the VPCs if any. We checked: (i) if the particle could be taken out without a major change in the meaning, (ii) if the particle expressed RESULT or could be replaced with a RESULT-Prepositional Phrase,[9] (iii) if a corresponding VPC consisting of the particle with the opposite

---

[8]While we do find prepositional phrasal verb constructions with *into*, we did not find any VPCs involving intransitive usages of *into*.

[9]RESULT is one of the argument roles identified in the TRIPS ontology. The argument roles signal different argument positions for predicates as well as have their own inferential import, some other examples are AGENT, AFFECTED, MANNER, LOCATION, and FIGURE.

polarity was also possible (e.g., *take in* vs. *take out*), (iv) if specific argument types (e.g., MANNER, RESULT, LOCATION, AFFECTED) were instantiated in the sentence, etc. In the rest of this section, we present the sense classes particles in compositional VPCs tend to fall into.

4.1 Sense classes for particles in VPCs

While particles may encode subtle nuances of meanings in each of their occurrences in (compositional) VPCs, they may also display some general senses across many VPCs. WordNet attempts to capture the nuances by storing each of the VPCs as a separate lexical item. However, this approach results in having as many sense categories as there are VPCs and we lose information about the common contributions made by the particles in VPC semantics which can be useful while producing semantic representation of sentences with new VPCs not stored in WordNet or another lexical resource. Hence, we focus on the general senses particles display across compositional VPCs.

We identified two general sense classes for the particles in compositional VPCs, namely DIRECTION and ASPECTUAL. The DIRECTION sense class has a number of subclasses, each instantiated by a specific directional particle, such as *away*, *down*, *in*, *off*, *on*, *out*, and *up* denoting a specific direction sense. For example, the directional particle *away* instantiates the subclass DIRECTION-AWAY in *I took one last look at the house and **walked away***. Similarly, the particle *out* in *My mom never **threw** it **out*** and the particle *up* in *The magic ketchup should sink when you squeeze the bottle and **float up** when you release it* instantiate subclasses DIRECTION-OUT and DIRECTION-UP respectively. The DIRECTION sense class particles assume the RESULT role in relation to the verb. For example, in the DIRECTION-AWAY example, the particle *away* denotes the RESULT of a *walking* event in terms of the direction the AFFECTED entity is walking to.

The ASPECTUAL sense class has two subclasses, namely COMPLETELY and CONTINUING, where the particle modifies the verb by providing aspectual information. In terms of the TRIPS semantic roles, the aspectual sense particles assume the MANNER role in relation to the verb. The COMPLETELY sense is used to express that the activity denoted by the verb is performed to the full extent or with thoroughness. Particles with COMPLETELY sense may also emphasize the telicity of an action. For example, the particle *out* in *He **sorted out** every scrap of manuscript, every map, and the native letters* emphasizes that each of the items mentioned were thoroughly and completely sorted. Particles *up*, *out*, and *down* are used more often than other particles to convey this aspectual sense.

The CONTINUING sense is used to emphasize the durative nature of the event denoted by the verb. For example, in *Day after day she **worked away** ...* the particle *away* conveys that the activity continued for a duration. The particles in this class may also convey an iterative sense when they are used with semelfactive verbs, e.g., note the use of the particle *away* in *Start your explorations here, **click away** all you want*. Usually particles *away* and *on* are used to convey this sense in VPCs.

These sense classes correspond to the VPC classes based on their compositionality types mentioned in §3. For example, the DIRECTION sense class is generally instantiated by the symmetrically compositional or LV-compositional VPCs. The ASPECTUAL sense class, on the other hand, is instantiated by the LP-compositional VPCs. Besides these two sense classes, there are a few other senses that particles may express in VPCs corresponding to the symmetrically compositional and LV-compositional VPCs. For example, the senses IN-WORKING-ORDER-VAL and NOT-IN-WORKING-ORDER-VAL mentioned in Table 2 and Table 4 are expressed by particles in LV-compositional VPCs. These senses are usually conveyed by the particles *up*, *down*, and *out*. Similarly, the sense DIS-APPEARANCE, conveyed by the particle *away*, may be taken to correspond to the particle in symmetrically compositional VPCs. Table 2 contains information about each of these senses and corresponding sense classes for the particles *up*, *down*, *away*, and *out*. Table 3 and Table 4 present generalizations corresponding to these senses and sense classes.

We did not get very high agreement on human annotations for the sense labels for the particles in VPCs from Test Set 1. The three annotators had full agreement in 20 cases (40% of the cases). 75% of these 20 cases had DIRECTION sense, 20% had COMPLETELY, and 5% IDIOMATIC. The disagreement cases involved labels such as DIRECTION vs. COMPLETELY, DIRECTION vs. IDIOMATIC, or COMPLETELY vs. IDIOMATIC. For a more reliable identification of senses for the particles in VPCs, one may adopt tests as in Figure 2.

We also explored the use of compositionality heuristics to determine the senses of particles in VPCs, the general sense labels being (DIRECTION, ASPECTUAL, and IDIOMATIC). For example, if heuristic 3 applied to a VPC, the particle was assumed to have an ASPECTUAL sense. If heuristic 7 applied, the VPC was taken to be noncompositional and hence with an IDIOMATIC sense. In 30% of the cases, a correct sense label was assigned unambiguously. In another 32% of the cases, multiple sense labels were identified and one of them was correct. In the rest of the 19 cases where the sense label based on the application of specific heuristics was incorrect, the DIRECTION sense was the most misclassified sense (57.89%), followed by the COMPLETELY sense (31.58%), followed by the IDIOMATIC sense (10.53%).

Table 2: Sense classes for particles *up*, *down*, *away*, and *out* in compositional VPCs.

Sense class Sense type	Particle	Example
DIRECTION:		
DIRECTION-UP	up	He took the carpet up.
DIRECTION-DOWN	down	He skied down.
DIRECTION-AWAY	away	He ran away.
DIRECTION-OUT	out	He cried out.
ASPECTUAL:		
COMPLETELY	up	Clean up the room!
	down	London nightclub closed down over fights with knives and bottles.
	out	He sorted out every scrap of manuscript, every map, and the native letters.
ASPECTUAL:		
CONTINUING	away	Night and day they hammered away, coming on like great waves. He scrubbed away at the floor.
IN-WORKING-ORDER-VAL	up	Bring the browser up!
NOT-IN-WORKING-	down	The computer went down again.
ORDER-VAL	out	The national electric grid went out.
DISAPPEARANCE	away	The echo died away. The music faded away.

Table 3: Generalizations about sense classes DIRECTION and ASPEC-
TUAL for particles in compositional VPCs.

Sense class: Sense type	Generalizations
DIRECTION:	
DIRECTION-UP DIRECTION-DOWN DIRECTION-AWAY DIRECTION-OUT	verb trajectory = + Particle relative to some scale/domain PP alternative is possible for the particles *up*, *down*, and *out*. The particle *away* can take a PP-location as its GROUND argument. The senses tend to appear with motion and movement-related verbs in the TRIPS ontology. Cases pass two tests: (1) Is there a change in physical location/on a scale? (2) Does the entity exist anywhere? DIRECTION-UP is also possible with verbs in the TRIPS ontology classes apply-force and acquire. DIRECTION-DOWN is also possible with verbs in the TRIPS ontology classes hitting. DIRECTION-AWAY is also possible with event-of-change verbs where grammatical resultative construction is used. DIRECTION-OUT is also possible with verbs of vocalization (and the verbs in the TRIPS ontology classes locution and manner-say) and some perception verbs.
ASPECTUAL:	
COMPLETELY	verb trajectory = - PP alternative not possible Default sense with the event-of-change verbs in the TRIPS ontology COMPLETELY for the particle *up* is also possible with verbs in the TRIPS ontology classes protecting, joining, acquire-by-action, herd, and arrange-text. COMPLETELY for the particle *down* is also possible with verbs in the TRIPS ontology classes change, consume, protecting, put, and pursue. COMPLETELY for the particle *out* is also possible with verbs in the TRIPS ontology classes evoke-tiredness, arranging, and put.
ASPECTUAL:	
CONTINUING	verb trajectory = - Usually with atelic verbs that have extended duration Iterative usage with semelfactive verbs PP alternative is not possible, i.e., particle cannot take NP object, but can take PP object in conative constructions. CONTINUING is also possible with verbs in the TRIPS ontology class event-of-action.

Table 4: Generalizations about senses IN-WORKING-ORDER-VAL, NOT-IN-WORKING-ORDER-VAL, and DISAPPEARANCE for particles in compositional VPCs. Note that NP* refers to the object argument of a transitive verb, e.g., *the carpet* in *he took the carpet up*, or the subject of an intransitive verb, e.g., *he* in *he skied down*.

| Sense class: Sense type	Generalizations
IN-WORKING-ORDER-VAL	NP* is phys-obj object-function = provides-service-up-down i.e., the entity is a device with some functionality. PP alternative not possible The verbs tend to be light/causal verbs.
NOT-IN-WORKING-ORDER-VAL	NP* is phys-obj object-function = provides-service-up-down i.e., the entity is a device with some functionality. PP alternative not possible NOT-IN-WORKING-ORDER-VAL is also possible for the particle *down* with verbs in the TRIPS ontology class event-of-undergoing-action. The verbs tend to be light/causal verbs.
DISAPPEARANCE	verb trajectory = - Usually with verbs of disappearance PP alternative not possible Cases fail two tests: (1) Is there a change in physical location/on a scale? (2) Does the entity exist anywhere? DISAPPEARANCE is also possible with verbs in the TRIPS ontology classes event-of-undergoing-action and change.
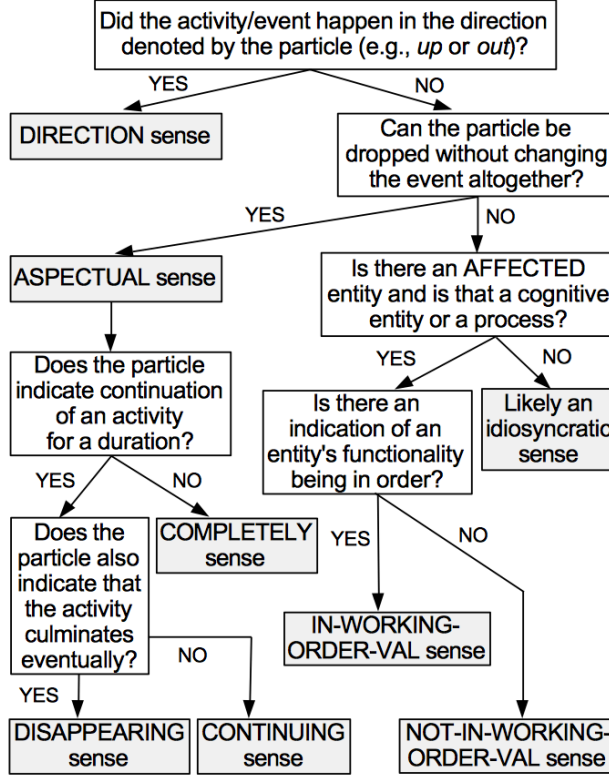
Figure 2: Tests to identify particle senses

5 Computing semantics of sentences with VPCs

For the task of interpreting sentences with VPCs, we first need to determine if the VPC is compositional or not. We use the heuristics mentioned in §3.2 to determine the compositionality of a VPC. For the compositional cases, we get the senses for the verb and the particle from the TRIPS ontology and/or Word-Net. The senses for the particles as well as relevant linguistic generalizations to identify these senses are encoded in the TRIPS lexicon and ontology. In this section, we briefly discuss some of the generalizations encoded in the ontology and demonstrate the process of computing the semantics of sentences containing compositional VPCs using three example sentences involving VPCs with different senses for the particle *up*: *She cleaned up her room*, *She pushed the ball up*, and *The network came up*. The logical forms (LFs) produced for these sentences using the TRIPS parser, a broad coverage deep semantic parser driven by the TRIPS ontology, are presented in Figure 3 to Figure 5.

ONT::PUSH

AGENT   RESULT

AFFECTED

She   *The ball*   ONT:DIRECTION-UP
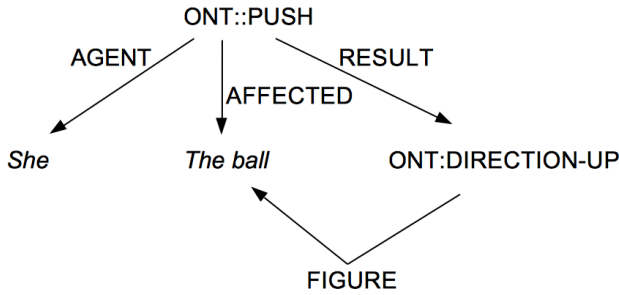
FIGURE

Figure 3: LF for the sentence: *She pushed the ball up.*

ONT::CLEAN

AGENT   MANNER   FIGURE

AFFECTED

She   *her room*   ONT::COMPLETELY
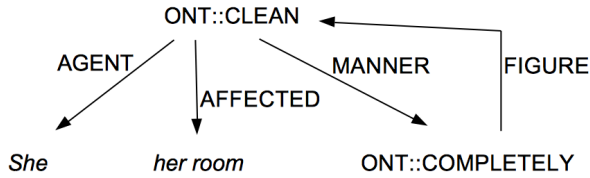
Figure 4: LF for the sentence: *She cleaned up her room.*

ONT::BECOME

AFFECTED   FORMAL

The network   ONT::IN-WORKING-ORDER
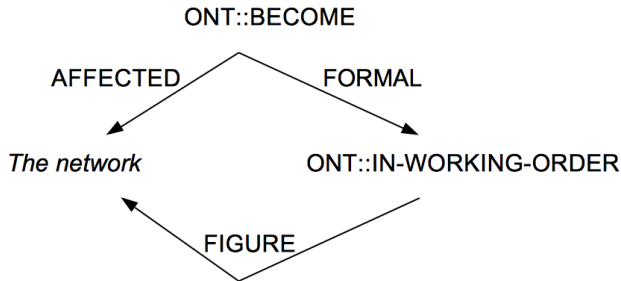
FIGURE

Figure 5: LF for the sentence: *The network came up.*

Particles in compositional VPCs can express the senses mentioned in Tables 2–4, and in Figure 2. This information is encoded in the TRIPS ontology by adding the ontology types corresponding to these senses in the particle's lexicon. For example, the lexical entry for the particle *up* lists sense ontology types ONT::DIRECTION-UP, ONT::COMPLETELY, and ONT::IN-WORKING-ORDER-VAL, used in Figure 3 to Figure 5 respectively, among other possible senses.[10]

---

[10]For a better idea of what information the lexical entries and semantic/ontology classes carry in the TRIPS lexicon and ontology, see http://www.cs.rochester.edu/research/cisd/projects/trips/lexicon/browse-ont-lex-ajax.html.

WordNet sense keys corresponding to the particle may be added in the ontology entries for these sense ontology types. For example, WordNet sense keys up%4:02:00 and up%4:02:05 are added in the entry for the sense ontology type ONT::DIRECTION-UP.

The senses that particles express in a VPC may depend on the verb type they combine with in the VPC. That is, a particle may convey the same sense when it appears with any of the verbs in a specific verb ontology class.[11] For example, particles tend to get DIRECTION sense with verbs of motion and CONTINUING sense with atelic verbs that have extended durations (e.g., activity type verbs). With verb ontology types corresponding to continuous-change verbs (which appear under change verbs in the TRIPS ontology), particles tend to get the COMPLETELY sense. Specific verb ontology classes are also identified for specific particles, e.g., particle *down* exhibits COMPLETELY sense with the verbs in the TRIPS ontology class ONT::PURSUE, as can be seen in *The internet* **tracked down** *this guy's stolen car (...)* and *A motorist* **chased down**, *slapped, and threatened a boy (...)*.

In addition, we observed an interesting fact that the particles *up* and *out* seem to be in complementary distribution with respect to various verb ontology classes for the COMPLETELY sense. That is, for the COMPLETELY sense, either *up* or *out* is used, but not both with verbs from a specific verb ontology class.[12] For example, with the verb ontology class ONT::ACQUIRE, *up* is used with COMPLETELY sense, *out* cannot be used with the verbs in this ontology class with the same sense. Note the COMPLETELY sense in the VPC *acquire up* in *Techstars has* **acquired up** *Global*, but we do not observe a VPC *acquire out* with the same sense. Similarly, with the verb ontology class ONT::EVOKE-TIREDNESS, *out* is used with the COMPLETELY sense, but *up* cannot be used. Note that we can say, *Someone's a bit* **tuckered out**, but not **tuckered up**.

Similarly there are generalizations observed for specific senses of particles corresponding to the semantic relation labels. For example, the verb takes a particle with an ASPECTUAL sense as its MANNER argument and the ASPECTUAL sense particle takes the verb as a FIGURE, as is observed in the LF in Figure 4. For the DIRECTION sense class particles, the verb assigns a RESULT argument

---

[11]We find that different verb ontology types that were distinguished for other reasons in the TRIPS ontology (Allen et al. 2007) also line up with the particle distinctions.

[12]This observation about the complementary distribution of usage between *up* and *out* may not be accidental. The Law of Differentiation (Paul 1890; Bréal 1900), and the Avoid Synonymy principle (Kiparsky 1983; Clark 1987) have been proposed in the lexico-semantic sphere which suggest that languages prefer not to have a given semantic slot be filled by two distinct lexical items.

Draft of October 22, 2018, 16:48

role instead of a MANNER role to the particle, as illustrated in Figure 3. The particle, on the other hand, assigns the FIGURE role to the AFFECTED entity as in Figure 3. For the IN-WORKING-ORDER-VAL sense, the particle takes the AF-FECTED or NEUTRAL entity as its FIGURE argument as is illustrated in Figure 5. Thus, we see that different semantic relations may be involved in sentences with different senses for the particles.

In order to get the correct semantic relations in different cases, we encode this information in the ontology. The sense ontology type ONT::COMPLETELY, used in Figure 4, for example, specifies for its FIGURE argument all the verb ontology types with which a particle gets this sense. This information is presented in the Generalizations column in Table 3. Note that under ONT::COMPLETELY, we list all the verb ontology types with which we get the COMPLETELY sense irrespective of the specific particles used.[13] Hence, ONT::COMPLETELY would specify for its FIGURE argument the ontology types ONT::PURSUE, ONT::ACQUIRE as well as ONT::EVOKE-TIREDNESS, for example.

Each of the verb ontology types mentioned above specifies which particles can take a specific semantic relation label. For example, the verb ontology type ONT::PURSUE would specify for its MANNER argument particle *down*, verb ontology type ONT::ACQUIRE would specify particle *up*, and verb ontology type ONT:EVOKE-TIREDNESS would specify particle *out*.

For the DIRECTION senses, the particle could be replaced with a RESULT-Prepositional Phrase (RESULT-PP) in the sentence. For example, the particle *off* in *The officer got **off** when he spotted an illegally parked car* can be replaced with a RESULT-PP *off his motorcycle* as in *The officer got **off his motorcycle** when he spotted an illegally parked car.*

Depending on the compliance or violations of constraints such as the ones described above, the parser assigns scores for various parse options involving various senses of the particle in a VPC. The parse with the highest score is selected as the semantic representation of the sentence involving the VPC.

In *She cleaned up her room* (Figure 4), the verb *clean* (ONT::CLEAN which appears under ONT::CHANGE-STATE in the ontology) is not among the list of relevant verb ontology types with which a DIRECTION-UP sense is licensed for the particle *up*. Additionally, a restriction on the verb argument for the DIRECTION-UP sense is that the argument have a semantic feature [+moveable] which is also violated in the given sentence, since *the room* is generally not a moveable entity. Hence, the parser assigns a low score to the parse which involves the

---

[13]Note that since the ontology is hierarchical, there is no need to list all the children ontology types as well if the parent ontology types are included.

DIRECTION-UP sense for the particle *up* in this sentence. The verb *push*, however, in the sentence in Figure 3, is a movement-related verb and the affected entity *ball* is a moveable object. Hence, the parser assigns a higher score to the parse that involves the DIRECTION-UP sense for the particle *up* in this sentence.

The IN-WORKING-ORDER-VAL sense requires restrictions on the verbs that they take cognitive entities, devices or processes as their AFFECTED arguments which provide service, i.e., they have some functionality associated with them. The AFFECTED argument for the verb *clean*, namely *the room*, in Figure 4 does not satisfy this restriction. Similarly, The AFFECTED argument for the verb *push*, namely *the ball*, in Figure 3 does not satisfy this restriction. Hence, the parser assigns a low score for the parses involving an IN-WORKING-ORDER-VAL sense for the particle *up* in the two sentences. On the other hand, the entity *The network* in the sentence in Figure 5 satisfies the restriction and hence the sentence in Figure 5 gets a higher score for the parse that contains IN-WORKING-ORDER-VAL sense for the particle *up*.

The constraints for the COMPLETELY sense of the particle *up* are satisfied for the sentence in Figure 4. The verb *clean* is among the set of verbs in the ontology type (ONT::CHANGE-STATE) with which the relevant particle has been identified in the ontology to get this sense. Hence, the parser assigns a higher score to the parse for the sentence with the COMPLETELY sense for the particle *up*. The other two sentences get a lower score for a parse with the COMPLETELY sense for the particle *up*.

Thus, as mentioned above, the parse with the highest score is selected as a semantic representation for each of the sentences involving the VPC. Hence, the sentence in Figure 3 has the DIRECTION-UP sense for the particle *up*, the sentence in Figure 4 has the COMPLETELY sense, and the sentence in Figure 5 has the IN-WORKING-ORDER-VAL sense for the particle *up*.

6 Conclusion

Identification of MWEs, such as verb-particle constructions, as well as the computation of their meaning is necessary for a broad coverage understanding. It is not plausible to hand enumerate all the possible combinations, although Word-Net is an admirable start. We have described an approach where the meaning of a wide range of VPCs is computed compositionally, with the advantage that VPCs not explicitly found in the lexicon can be both identified and semantically interpreted. To accomplish this, we identified the core senses of particles that have broad application across verb classes. This information is used while

building computational lexica. We encoded the generalizations corresponding to various senses of particles in the TRIPS ontology and used them to identify these senses. We found that the generalizations corresponding to grammatical/semantic/ontological information help us identify appropriate senses of the particle. The procedure adopted enables compositional parsing by helping differentiate between particle senses, which is then used to obtain full semantic representations of sentences.

While we have outlined our approach to compute the semantics of sentences with VPCs in English using resources such as the TRIPS lexicon and ontology and WordNet, a similar approach can be adopted to incorporate more lexical or other linguistic resources to further improve semantic parsing of sentences involving VPCs in English. Also, such an approach can be adopted for semantic parsing of sentences involving other MWEs, as well as for developing semantic parsing capabilities for other languages using the resources available for those languages.

Abbreviations

LF	logical form
LP-compositional	light particle compositional
LV-compositional	light verb compositional
MWE	multiword expression
RESULT-PP	RESULT-prepositional phrase
VPC	verb-particle construction

References

Allen, James F., Myroslava Dzikovska, Mehdi Manshadi & Mary Swift. 2007. Deep linguistic processing for spoken dialogue systems. In *Proceedings of the ACL 2007 Workshop on Deep Linguistic Processing*, 49–56. Association for Computational Linguistics. June 28, 2007.

Allen, James F., Mary Swift & Will de Beaumont. 2008. Deep semantic analysis of text. In *Symposium on semantics in systems for text processing (STEP)*, 343–354.

Allen, James F. & Choh Man Teng. 2017. Broad coverage, domain-generic deep semantic parsing. In *Proceedings of the AAAI Spring Symposium, Computational Construction Grammar and Natural Language Understanding 2017*, 108–115.

Baldwin, Timothy, Colin Bannard, Takaaki Tanaka & Dominic Widdows. 2003. An empirical model of multiword expression decomposability. In *Proceedings of the ACL 2003 Workshop on Multiword Expressions: Analysis, Acquisition and Treatment - Volume 18* (MWE '03), 89–96. Stroudsburg, PA, USA: Association for Computational Linguistics. DOI:10.3115/1119282.1119294

Bannard, Colin James. 2002. *Statistical techniques for automatically inferring the semantics of verb-particle constructions.* Tech. rep. University of Liverpool.

Bannard, Colin James & Timothy Baldwin. 2003. Distributional models of preposition semantics. In *Proceedings of the ACL-SIGSEM Workshop on the Linguistic Dimensions of Prepositions and their Use in Computational Linguistics Formalisms and Applications*, 169–180. Association for Computational Linguistics. http://lingo.stanford.edu/pubs/tbaldwin/sigsemprep2003-prepsem.pdf.

Bannard, Colin, Timothy Baldwin & Alex Lascarides. 2003. A statistical approach to the semantics of verb-particles. In *Proceedings of the ACL 2003 Workshop on Multiword Expressions: Analysis, Acquisition and Treatment - Volume 18* (MWE '03), 65–72. Stroudsburg, PA, USA: Association for Computational Linguistics. DOI:10.3115/1119282.1119291

Bhatia, Archna, Choh Man Teng & James F. Allen. 2017. Compositionality in verb-particle constructions. In *Proceedings of the 13th Workshop on Multiword Expressions* (MWE '17), 139–148. Association for Computational Linguistics. http://www.aclweb.org/anthology/W17-1719.

Bréal, Michel. 1900. *Semantics.* Trans. by Mrs. H. Cust. New York: Henry Holt.

Chen, Ping. 1986. Discourse and particle movement in English. *Studies in Language* 10. 79–95.

Clark, Eve V. 1987. The principle of contrast. In Brian MacWhinney (ed.), *Mechanisms of language acquisition*, 1–33. New York: Academic Press.

Cohen, Jacob. 1960. A coefficient of agreement for nominal scales. *Educational and Psychological Measurement* 20. 37–46.

Cook, Paul & Suzanne Stevenson. 2006. Classifying particle semantics in English verb-particle constructions. In *Proceedings of the Workshop on Multiword Expressions: Identifying and Exploiting Underlying Properties* (MWE '06), 45–53. Association for Computational Linguistics. July 23, 2006.

Dehé, Nicole. 2002. *Particle verbs in English: Syntax, information structure and intonation.* Amsterdam: John Benjamins.

Fellbaum, Christiane. 1998. *WordNet: An electronic lexical database.* Cambridge, MA: MIT Press.

Fleiss, Jacob L. 1971. Measuring nominal scale agreement among many raters. *Psychological Bulletin* 76(5). 378–382.

Fraser, Bruce. 1976. *The verb-particle combination in English*. New York: Academic Press.

Ishizaki, Yasuaki. 2010. Some notes on the grammaticalization of *away*. In Hirozo Nakano, Masayuki Ohkado, Tomoyuki Tanaka, Tomohiro Yanagi & Azusa Yokogishi (eds.), *Synchronic and diachronic approaches to the study of language: A collection of papers dedicated to the memory of Professor Masachiyo Amano*, 71–83. Tokyo: Eichosha.

Ishizaki, Yasuaki. 2012. A usage-based analysis of phrasal verbs in early and late modern English. *English Language and Linguistics* 16(2). 241–260. DOI:10.1017/S1360674312000020

Jackendoff, Ray. 1997. *The architecture of the language faculty*. Cambridge, MA: MIT Press.

Jackendoff, Ray. 2002. English particle constructions, the lexicon, and the autonomy of syntax. In Nicole Dehé, Ray Jackendoff, Andrew McIntyre & Silke Urban (eds.), *Verb-Particle Explorations* (Interface Explorations [IE]), 67–94. New York: Mouton de Gruyter.

Johnson, Christopher & Charles J. Fillmore. 2000. The FrameNet tagset for frame-semantic and syntactic coding of predicate-argument structure. In *Proceedings of the First Conference on North American Chapter of the Association for Computational Linguistics*, 56–62. Morgan Kaufmann Publishers Inc.

Kiparsky, Paul. 1983. Word-formation and the lexicon. In Frances Ingemman (ed.), *Proceedings of the 1982 Mid-America Linguistics Conference*, 47–78. University of Kansas, Dept. of Linguistics.

McCarthy, Diana, Bill Keller & John Carroll. 2003. Detecting a continuum of compositionality in phrasal verbs. In *Proceedings of the ACL 2003 Workshop on Multiword Expressions: Analysis, Acquisition and Treatment - Volume 18* (MWE '03), 73–80. Stroudsburg, PA, USA: Association for Computational Linguistics. DOI:10.3115/1119282.1119292

Miller, George A. 1995. WordNet: A lexical database for English. *Communications of the ACM* 38(11). 39–41.

Moreno-Ortiz, Antonio, Chantal Pérez-Hernández & M. ángeles Del-Olmo. 2013. Managing multiword expressions in a lexicon-based sentiment analysis system for Spanish. In *Proceedings of the 9th Workshop on Multiword Expressions* (MWE '13), 1–10. Association for Computational Linguistics.

Niinuma, Fumikazu. 2014. Grammaticalization of the particle *away* in English: A cartographic approach. *Interdisciplinary Information Sciences* 20(2). 163–188.

O'Dowd, Elizabeth M. 1998. *Prepositions and particles in English: A discourse-functional account*. Oxford: Oxford University Press.

Patrick, Jon & Jeremy Fletcher. 2004. Differentiating types of verb particle constructions. In *Proceedings of Australasian Language Technology Workshop 2004* (ALTW 2004). http://aclweb.org/anthology-new/U/U04/U04-1022.pdf. December 8, 2004.

Paul, Hermann. 1890. *Principles of the history of language.* Trans. by H. A. Strong from the 2nd German Edition. Reprinted College Park, MD. 1970: McGrat.

Tyler, Andrea & Vyvyan Evans. 2003. *The semantics of English prepositions: Spatial scenes, embodied meaning, and cognition.* New York: Cambridge University Press.

Villavicencio, Aline. 2003. Verb-particle constructions and lexical resources. In *Proceedings of the ACL 2003 Workshop on Multiword Expressions: Analysis, Acquisition and Treatment* (MWE 2003)- *Volume 18*, 57–64. Association for Computational Linguistics. DOI:10.3115/1119282.1119290

Villavicencio, Aline. 2006. Verb-particle constructions in the World Wide Web. In Patrick Saint-Dizier (ed.), *Syntax and semantics of prepositions*, 115–130. The Netherlands: Springer.

Vincze, Veronika, István Nagy T. & Gábor Berend. 2011. Multiword expressions and named entities in the Wiki50 corpus. In *Proceedings of the International Conference Recent Advances in Natural Language Processing 2011*, 289–295. RANLP 2011 Organising Committee. http://aclweb.org/anthology/R11-1040.

Draft of October 22, 2018, 16:48

Did you like this book?

This book was brought to you for free

Please help us in providing free access to linguistic research worldwide. Visit http://www.langsci-press.org/donate to provide financial support or register as a community proofreader or typesetter at http://www.langsci-press.org/register.

language
science
press

Multiword expressions at length and in depth

The annual workshop on multiword expressions takes place since 2001 in conjunction with major computational linguistics conferences and attracts the attention of an ever-growing community working on a variety of languages, linguistic phenomena and related computational processing issues. MWE 2017 took place in Valencia, Spain, and represented a vibrant panorama of the current research landscape on the computational treatment of multiword expressions, featuring many high-quality submissions. Furthermore, MWE 2017 included the first shared task on multilingual identification of verbal multiword expressions. The shared task, with extended communal work, has developed important multilingual resources and mobilised several research groups in computational linguistics worldwide.

This book contains extended versions of selected papers from the workshop. Authors worked hard to include detailed explanations, broader and deeper analyses, and new exciting results, which were thoroughly reviewed by an internationally renowned committee. We hope that this distinctly joint effort will provide a meaningful and useful snapshot of the multilingual state of the art in multiword expressions modelling and processing, and will be a point of reference for future work.