Compositionality in Verb-Particle Constructions

Archna Bhatia[1], **Choh Man Teng**[2], **and James F. Allen**[2,3]

[1]Florida Institute for Human and Machine Cognition, 15 SE Osceola Ave, Ocala, FL 34471
[2]Florida Institute for Human and Machine Cognition, 40 S Alcaniz St, Pensacola, FL 32502
[3]Department of Computer Science, University of Rochester, Rochester, NY 14627
{abhatia,cmteng,jallen}@ihmc.us

Abstract

We are developing a broad-coverage deep semantic lexicon for a system that parses sentences into a logical form expressed in a rich ontology that supports reasoning. In this paper we look at verb-particle constructions (VPCs), and the extent to which they can be treated compositionally vs idiomatically. First we distinguish between the different types of VPCs based on their compositionality and then present a set of heuristics for classifying specific instances as compositional or not. We then identify a small set of general sense classes for particles when used compositionally and discuss the resulting lexical representations that are being added to the lexicon. By treating VPCs as compositional whenever possible, we attain broad coverage in a compact way, and also enable interpretations of novel VPC usages not explicitly present in the lexicon.

1 Introduction

Toward the goal of Natural Language Understanding of full interpretation of a text fragment (or a sentence), we want to produce a good semantic representation of the sentence. This involves combining rich grammatical information with information about specific lexical items in the sentence, such as word senses, among other things. Since 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, it is important to get a good interpretation of MWEs.

For this paper, we focus on a specific type of MWEs, namely verb-particle constructions (VPCs). These 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 in a lexicon is possible but not ideal as, for example, some VPCs may be interpretable compositionally, i.e., the verb and the particle contribute their simplex meanings, e.g. *fly up*. Other compositional VPCs include cases such as *finish up* and *made away* for which either the verb or the particle, respectively, seems to contribute its simplex meaning (Bannard et al., 2003).[2] However, other VPCs indeed are noncompositional and require special interpretation, and hence need to be added into the lexicon, e.g., *bake off* 'contest' and *egg on* 'urge someone for an action that might not be a good idea'.

For an interpretation of the compositional types above, we need to determine the best senses for the verb and the particle in the VPCs. There are many lexical resources for an inventory of senses for verbs, such as WordNet (Miller, 1995), (Fellbaum, 1998) and VerbNet (Kipper-Schuler, 2005). But there is not much for the particles except for a few attempts at the semantics for a few particles, such as *up* (Cook and Stevenson, 2006) and *out* (Tyler and Evans, 2003). Our investigation of hundreds of VPCs has shown that the semantics of particles is also important, as can also be gathered from others' proposals for similar classifications of VPCs as mentioned above involving VPC types where particles contribute to the meaning, see Section 3 for details. Particles are not just the vacuous entities structurally required by the verbs in VPCs, they also have their own semantics which is found to be general across verbs

[1]Note we focus on the particle usage in this paper, 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 Section 4.1.

[2]However, refer to Section 3 for our take on such cases.

Proceedings of the 13th Workshop on Multiword Expressions (MWE 2017), pages 139–148,
Valencia, Spain, April 4.

in specific verb classes. For example, 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). Hence, in this paper, we provide a set of senses that particles in VPCs display across many verbs in a verb class.

To make use of these senses, we encode semantics of particles in an ontology, namely TRIPS (Allen et al., 2007) LF ontology which is designed to be linguistically informed.[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 and Fillmore, 2000). The ontology uses a rich semantic feature set, the features used are an extended version of EuroWordNet (Vossen, 1997). Unlike WordNet, the TRIPS ontology does not attempt to capture all possible word senses but rather focuses on the level of abstraction that affects linguistic processing. We use TRIPS, a broad coverage deep semantic parser (driven by the ontology) to combine semantic, ontological and grammatical information to produce semantic representation. For a more detailed overview of the TRIPS system, refer Allen & Teng (2017) and Allen et al. (2008).[4]

The paper is organized as follows: Previous work on VPCs is discussed in Section 2. In Section 3, classification of VPCs is discussed based on their compositionality. A set of heuristics are presented to identify different classes of VPCs in Section 3.1. In Section 4, we discuss the semantics of particles in VPCs. An inventory of general sense classes for particles used in VPCs is provided in Section 4.1. In Section 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. This also includes a discussion of phenomena we cannot handle currently. In Section 6, we demonstrate the procedure to compute meaning of sentences involving compositional VPCs. Finally, in Section 7, we present our conclusions.

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

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

2 Related work

A lot of computational literature on VPCs focuses on identification or extraction of VPCs, or on 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.

Vincze (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 and Fletcher (2004) also have a similar approach but focus on automatic classification of different types of compositionality. Unlike our work, in all these works, the focus is on compositionality only, not on actual senses of the particles.

Cook and Stevenson (2006) discuss various senses for the particle *up* in cognitive grammar framework and 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, e.g. 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 also based on the similar assumption that there are generalizations, such as combinations of particles with specific verb classes or ontological classes result in specific sense classes for the particles. 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 works mentioned above in the following respect: we emphasize on building complete semantic representations of the sentences, not just on particles' semantics or just classification of VPCs. Similar to our criteria for compositionality, McCarthy et al. (2003), Baldwin et al. (2003), 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, we present our heuristics for classification of VPCs based on WordNet and discuss how we compute the semantics of the compositional classes.

3 Classification of VPCs

VPCs have often been classified in terms of their compositionality/decomposability (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 or not), the classes following somewhere between the fully compositional and fully idiomatic VPCs, e.g., see Fraser (1976), Chen (1986), O'Dowd (1998), Dehé (2002) and Jackendoff (2002).[5]

In Figure 1, we present our classification of VPCs which also mainly consists of two types, the compositional and the noncompositional VPCs. We further identify the compositional VPCs into three subtypes, the symmetrically compositional VPCs, the light particle compositional VPCs (LP-compositional VPCs) and the light verb compositional VPCs (LV-compositional VPCs). The symmetrically compositional VPCs refer to the VPCs where both the constituents, the verb and the particle, contribute their lexical-semantic content. For example, in *The plane **flew up** in no time*, 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 meaning of the VPC *fly up*. 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[6] than the other constituent which adds its regular lexical-semantic content.

1. **Compositional VPCs:**
   (a) **Symmetrically compositional:** Both verb and the particle contribute their simplex meanings.
   *The plane **flew up** in no time.*
   (b) **Light particle compositional (LP-compositional):** Verb contributes most of the semantic content. Particle contributes aspectual information.
   *Susan **finished up** her paper.*
   (c) **Light verb compositional (LV-compositional):** Particle contributes most of the semantic content. The verb is generally a light verb contributing a bleached meaning such as BECOME or CAUSE etc.
   *The thief **made away** with the cash.*
2. **Noncompositional VPCs:**
   (a) **Noncompositional with certain generalizations:** Neither verb nor particle contribute their literal senses but certain generalizations are involved in interpretation of the VPCs.
   *She **took up** photography/swimming [activities]. vs. She **took up** her position [responsibility/position].*
   (b) **Idiosyncratic Noncompositional:** Idiomatic usages
   *John wouldn't have done the dangerous experiment if his brother hadn't **egged** him **on**.*

Figure 1: Classification of VPCs

The LP-compositional VPCs involve particles which, instead of contributing a preposition like lexical semantic content, contribute aspectual information to the VPC. For example, in *Susan **finished up** her paper*, the verb *finish* contributes its

[5] Based on Hawkins' (2000) classification, Lohse et al (2004) provide another classification of VPCs, also related to compositionality, in terms of whether the constituents of a VPC are independently processable or if one or both of them are dependent on the other for appropriate lexical-semantic content. For example, in *They turned off the lights*, how *off* is interpreted is independent of it appearing in the VPC, note this sentence entails *The lights are off*, however the verb is not independent of the particle in the VPC for its semantic content, the sentence does not entail *They turned the lights*.
Their categories can also be largely mapped to the compositional and the noncompositional cases.

[6] 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.

regular lexical content (e.g., in WordNet, sense finish%2:30:02), however, the particle *up*, instead of contributing its regular lexical-semantic content (e.g., WordNet sense up%4:02:00) adds aspectual information that the action was completed (i.e., the *Completely* sense in our sense inventory). See Section 4.1 for the specific senses of particles.

Similarly, the LV-compositional VPCs involve particles with their regular lexical-semantic content but have light verbs which carry bleached meaning than the regular verbs, e.g., CAUSE, BECOME, etc. 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 Section 6.

The noncompositional VPCs also seem to have at least two subtypes based on whether their interpretation involves certain generalizations or if it is completely idiosyncratic. However, for the rest of this paper, we focus on the compositional VPCs.

3.1 Heuristics for compositionality of VPCs

As a first step toward interpretation of VPCs, we need to determine whether a given VPC is compositional or not. For this task, 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 64188 multi-words. Heuristics 1-7 below are used to identify compositional VPCs, whereas heuristic 8 indicates a noncompositional VPC.[7]

1. If the verb is among the list of light verbs, and WordNet does not have an entry for the VPC, it most likely 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. If a VPC exists and WordNet has an entry for the verb as well as for the particle but no entry for the VPC, VPC is (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, *sort out* (**sort_out%2:31:00**) and *sort* (**sort%2:31:00**) both appear in the same synset in WordNet.

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

5. If WordNet has the verb in the definition in the synset where VPC appears, 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 particle replaced with another particle in the same synset, VPC is either symmetrically compositional or LP-compositional (with the two particles in the same sense class). For example, *pull up* (**pull_up%2:35:00**) as well as *pull out* (**pull_out%2:35:00**) are in the same synset. In these VPCs, the particles *up* as well as *out* have the same general sense *Direction* (see Section 4.1 for an inventory of particle sense classes).

7. If WordNet has the relevant VPC as well as another VPC with the verb replaced with another verb in the same synset, 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.

8. If none of the above are true and the VPC in WordNet does not have any other item in its synset, the VPC is likely idiomatic. For example, the idiomatic VPC *catch up* (**catch_up%2:38:00**) does not have any other item in its synset.

[7] Note that we do not claim that these heuristics cover the VPCs exhaustively.

3.2 An evaluation of heuristics for compositionality of VPCs

We conducted an evaluation of the heuristics 3-8 which is described as follows. From among all the VPCs for which WordNet has an entry, we automatically extracted 25 random VPCs such that each of the 12 particles (that we investigated, see Section 4) was represented in the extracted VPCs. These test VPCs were manually annotated by three annotators for the compositionality labels, *Compositional* and *Noncompositional*. Since a VPC may have both compositional and noncompositional usages in different contexts, we restricted assignment of the annotation label for a specific VPC to only one label by considering the first synset/definition each of the VPCs had in WordNet. In case of disagreement among the three annotations, the annotators discussed reasons for their decisions and arrived at a consensus to create the Gold annotations for the VPCs. One of the VPCs was dropped from the test set as the annotators could not reconcile with respect to the VPC.

A python implementation of the heuristics was applied to the remaining 24 test VPCs. Like the manual annotations mentioned above, for the heuristics also, only those annotations were considered which were based on the first synset/definition of the VPC in WordNet. The VPCs that heuristics 3-7 identified as representative of their category were annotated as *Compositional*, whereas the VPCs identified by heuristic 8 were annotated as *Noncompositional*. These annotations were tested against the Gold annotations for the VPCs.

As mentioned earlier, our heuristics do not cover all the VPCs. Out of the 24 VPCs, the heuristics did not assign a label to four VPCs. Also additional two VPCs had to be disregarded due to assignment of labels to them based on synsets/definitions other than the first synset/definition in WordNet. For the remaining 18 VPCs, the heuristics achieved an overall accuracy of 72%. For compositional cases specifically, the heuristics got 82% correct labels, and for the noncompositional cases, the heuristics achieved an accuracy of 57%.

One of the cases that the heuristics misidentified, namely *fly by*, was merely due to the current implementation of the heuristic not involving inflectional variations of the verb. Note WordNet definition includes the verb *fly* but in its inflected form *flying*. The heuristic 5 could capture it if the implementation is refined to cover inflected forms of verbs.

Finally, note heuristics 1 and 2 could not be evaluated using the same procedure by extracting VPCs from WordNet randomly since heuristics 1 and 2 identify VPCs that are not included in WordNet.

4 Semantics of particles in VPCs

As mentioned in Section 3, particles contribute to the overall semantics of compositional VPCs. In order to study the contribution of particles in VPCs, we conducted an investigation of VPCs consisting of verbs in the ontology class ONT::EVENT-OF-CAUSATION in the TRIPS ontology. Currently, there are 1383 words with verb senses in this class (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 were possible as VPCs): *across*, *away*, *by*, *down*, *in*, *into*, *off*, *on*, *out*, *over*, *through*, and *up*. We searched for examples for each of the combinations using Google and manually went through each of the examples to test various things. For example, we checked if any of the verb or the particle contributed to the overall meaning of the VPC, identified the senses particles had in the VPCs if any, checked if the particle could be taken out without a major change in meaning, if the particle expressed RESULT or could be replaced with a RESULT-Prepositional Phrase,[8] if a corresponding VPC consisting of the particle with the opposite polarity was also possible, e.g., *take in* vs *take out*, if specific argument types, e.g., MANNER, RESULT, LOCATION, AFFECTED etc 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, on the one hand, particles may encode subtle nuances of meanings in each of their occurrences in (compositional) VPCs, on the other hand, they may display some general senses across many VPCs. WordNet attempts to capture the nuances by storing each of the VPCs as a separate

[8]RESULT is one of the argument roles identified in 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.

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 VPCs.

We identified three sense classes for the particles in compositional VPCs, namely *Direction*, *Ready/Active* and *Aspectual*, as illustrated in Figure 2. These sense classes also correspond to the VPC classes based on their compositionality mentioned in Section 3. For example, the *Direction* sense class is generally instantiated by the symmetrically compositional and LV-compositional VPCs. The *Ready/Active* sense class is instantiated by the LV-compositional VPCs, and the *Aspectual* sense class by the LP-compositional VPCs.

1. **Direction:**
   (a) **Away:** *Can Modi unlock $1 trillion worth of gold stashed AWAY in India's lockers?*
   (b) **Out:** *My mom never threw it OUT.*
   (c) **Up:** *The magic ketchup should sink when you squeeze the bottle and float UP when you release it.*
2. **Ready/Active [+/-]:**
   (a) **Ready:** *That won't take DOWN "the internet" though, just DNS resolution.*
   (b) **Active:** *Baby was having a good sleep and mom woke him UP.*
3. **Aspectual:**
   (a) **Completely:** *He sorted OUT every scrap of manuscript, every map, and the native letters, he looked THROUGH the files.*
   (b) **Continuing:** *Day after day she worked AWAY remaking the old Granville house into a home.*
   (c) **Starting:** *Ask AWAY the question.*

Figure 2: Sense classes for particles in VPCs

The *Direction* sense class has a number of subclasses, each instantiated by a specific directional particle, such as *away, down, in, into, off, on, out, up* denoting a specific direction sense.[9]

The *Ready/Active* sense class also is a broad class of senses for the particles ranging from usages such as *Take DOWN the internet* to *Wake him UP*. We consider these VPC usages compositional since the particles display the same senses independently of the VPC usages. For example, one could say *The network is DOWN* where the particle *down* appears outside of a VPC with the same *Ready* sense as in the VPC usage *Take DOWN the internet*. Similarly, for *Active* sense, compare *I'm UP* with *Wake him UP*. More examples for this class of senses include: *Bring UP the internet/browser*, *Set UP an expertiment*, *Get UP*, *He had passed OUT from an apparent drug overdose* and *Turn ON/OFF the switch*. The common theme across these senses seems to be that these usages involve as AFFECTED arguments cognitive entities or processes/machines which may become more ready/active or less ready/active.[10]

The *Aspectual* sense class has three subclasses, namely *Completely*, *Continuing*, and *Starting*, where the particle modifies the verb by providing aspectual information.

We can employ certain heuristics to identify some of these particle senses. Since, in the symmetrically compositional and LV-compositional VPCs, particles contribute significantly in the lexical-semantic content of the VPCs, if they are removed from the construction, part of the meaning is also lost or the meaning changes drastically, as can be observed in the case of the *Direction* sense particle *out* in the following: *Then I can move OUT* vs. *Then I can move*, and in the case of the *Ready/Active* sense particle *Down* in the following: *That won't take DOWN "the internet" though* vs. *That won't take "the internet" though*.

Also, if a particle has a *Direction* sense, then replacing the particle with another directional particle should result in a VPC with just a change in the direction in the sense, e.g., *pull UP-/DOWN the screen*. The directional particles share their senses with the corresponding prepositional usages.[11] Hence, the directional particles

[9] All of the sense subclasses of *Direction* are not illustrated in Figure 2 to avoid redundancy.

[10] Even though we have identified *Ready* and *Active* as two separate subclasses for this class of senses, there seems to be more grey area for it to be difficult to always distinguish between the two subclasses.

[11] The difference seems to be that in the directional parti-

are replaceable with corresponding Prepositional Phrases (directional-PPs). For example, the particle *down* in *I walked DOWN* can be replaced with a corresponding directional-PP, as in *I walked DOWN THE STREET*.

Similarly, for LP-compositional cases, which include all the *Aspectual* sense classes, we check if the particle can be dropped without a major change in meaning in VPCs. For example, for *Completely* sense, the particle seems to enhance/emphasize the meaning of the verb or indicates completion of the activity denoted by the verb and can be dropped without a major change in meaning, e.g., *clean (UP) the room*, *EC is preparing to arrange (UP) elections in party lines as well*, *Techstars has acquired (UP) Global*. Also, the particle can generally be replaced with MANNER adverbials *completely* and *thoroughly*. For the other two senses also, there is only a slight loss of aspectual information in the VPC when the particle is dropped.

5 Building the computational lexicon (for semantic parsing of VPCs)

In this section, we discuss our findings in regard to the above mentioned sense classes. We find that there are complicated interactions between the verb ontology types and particles as well as arguments of the VPCs. We first present some of these interactions and discuss how corresponding information is encoded in the TRIPS lexicon. This is followed by the interactions which cannot be encoded in the ontology currently and are left for future work.

Particles can express one or more of the senses listed in Figure 2 in different VPCs. This information is encoded in TRIPS ontology by adding 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, ONT::COMPLETELY and ONT::READY among other possible senses.[12] Simultaneously, WordNet sense keys corresponding to the particle *up* may be added in the ontology entries for these sense ontology types.

cle usage, there is an implicit argument which is explicitly present in the prepositional usage.

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

Starting with the observation that the particle's sense may depend on the verb it combines with in a VPC, further generalizations are possible. Our investigation demonstrated that the sense of the particle in a VPC may be conditioned by the type of verb it appears with (rather than just a single verb) in most of the cases. That is, particles may convey the same sense when they appear with any of the verbs in a specific verb ontology class.[13] For example, the 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 particles *up* and *out* seem to be in complementary distribution with respect to various verb ontology classes for the *Completely* sense. That is, for *Completely* sense, either *up* or *out* is used but not both with verbs from a specific verb ontology class.[14] For example, with 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. Notice *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 *Completely* sense, but *up* cannot be used. Notice *Someone's a bit **tuckered** OUT* but not ***tuckered** UP*.

There are certain other 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.

Such information is encoded in the ontology as restrictions on the relevant arguments for the relevant sense ontology types as well as relevant verb ontology types. For simplification for demonstration, we use the three example cases men-

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

[14]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 to not have a given semantic slot be filled by two distinct lexical items.

tioned above and show how such information is encoded in the ontology. First of all, as mentioned above, the lexical entry for the particle lists the senses it can convey. Hence, *down*, *up* and *out* would include ONT::COMPLETELY in their lexical entry. Also the entry for the sense ontology type ONT::COMPLETELY would include WordNet sense keys for particles *down*, *up* and *out*. The sense ontology type ONT::COMPLETELY specifies for its FIGURE argument all the verb ontology types with which a particle gets this sense. Note here we list all the verb ontology types with which we get the *Completely* sense irrespective of the specific particles with which we get the sense.[15] Hence, ONT::COMPLETELY would specify for its FIGURE argument ontology types ONT::PURSUE, ONT::ACQUIRE as well as ONT::EVOKE-TIREDNESS. The restriction with regard to specific particle is captured in the verb ontology type. Each of the verb ontology types specify for their MANNER argument all the particles that can take that role (i.e., they can get *Completely* sense). Hence, 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*.

Similar generalizations are available for various Direction senses and a similar approach is taken to encode corresponding information. The main difference lies in the semantic roles, e.g., for *Direction* sense class particles, the verb assigns a RESULT argument role instead of the MANNER argument role and the verb ontology types specify a less restricted set of particles for the RESULT argument (since many of the direction particles can constitute VPCs with the verbs in a specific verb ontology class). One of the classic examples for VPCs with direction sense particles is with verb ontology types corresponding to motion verbs.

5.1 Difficult cases

We describe below a few interactions between the verb ontology classes, particles, their senses and the verb arguments which the ontology does not have a way to handle currently. We leave these for future work.

[15] However, note 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.

We observe that the object (possibly the AFFECTED or AFFECTED-RESULT argument)[16] of the VPC may have an impact on the sense a particle gets. For example, in *I cleaned OUT **the desk***, the particle *out* is interpreted as having *Completely* sense whereas in *I cleaned OUT **the dirt***, it seems to have the *Direction* sense.

The order of the particle and the object may also affect the interpretation the particle gets. For example, in *help OUT **a friend***, the particle only gets the *Completely* sense. But in the reverse order for the particle and the object, e.g., in *help **a friend** OUT* as in "help a friend out of a difficult/unsafe situation", *Direction* sense is also possible.[17]

While ontology can specify semantic features for the verbal arguments for correct assignment of semantic roles to them, it cannot currently restrict senses that the particles may get in VPCs based on the semantic features of the verbal arguments. Similarly, the link between the argument position and the particle sense cannot be handled currently.

6 Procedure to compute meaning 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 heuristics mentioned in Section 3.1 to determine the compositionality of VPC. For the compositional cases, we get the senses for the verb and the particle from the ontology and/or WordNet. In the rest of this section, we walk through the process of computing the semantics of a sentence containing a compositional VPC using a broad coverage deep semantic parser driven by ontology.

Let's say the sentence we want to interpret is *She cleaned up her room*. The sentence involves the VPC *clean up* with the verb *clean* and the particle *up*. Let's say, the particle *up* has the following senses encoded in the ontology: *Direction*, *Completely*, and *Ready*. As mentioned in Section 5, there are certain constraints on verb ontology types (as well as verb/VPC arguments) for the parser to pick one of these senses of the particle when co-occurring with the verb. Depending on

[16] AFFECTED-RESULT is a semantic role our ontology uses for entities that undergo a change at the end of the event. The AFFECTED role is used for entities that changed over the course of the event in some way.

[17] In fact, this seems to be a relatively general pattern as is pointed out by Fraser (1976) that a directional adverbial tends to follow the construction.

compliance or violations of all such constraints, the parser assigns scores for various parse options involving these senses. The parse with the highest score is selected as a semantic representation of the sentence involving the VPC.

In the sentence *She cleaned up her room*, 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* sense is licensed for the particle *up*. Additionally, a restriction on the verb argument for the *Direction* sense is that the argument have a semantic feature [+moveable] which is also violated in the given sentence, *the room* is generally not a moveable entity. Hence, the parser assigns a low score to the parse which involves the *Direction* sense for the particle *up* in this sentence.

The *Ready/Active* sense requires restrictions on the verbs that they take cognitive entities or processes as their AFFECTED arguments. The AFFECTED argument for the verb *clean*, namely *the room*, does not satisfy this restriction. Hence, the parser assigns a low score for the parse involving a *Ready/Active* sense for the particle *up* in the given sentence.

The constraints for the *Completely* sense of the particle *up* are satisfied for this sentence, 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*. Among the three parses involving each of the above-mentioned senses of the particle, since the parse with the *Completely* sense gets the highest score, the parse is selected as the semantic representation of the sentence.

7 Conclusion

In order to attain broad coverage understanding, a system need not only identify multi-word expressions such as verb-particle constructions, but must compute their meaning. It is not plausible to hand enumerate all the possible combinations, although WordNet is an admirable start. We have described an approach where the meaning of a wide range of VPCs are computed compositionally, with the large 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 lexicons. We also discussed some difficult cases involving interesting interactions between verb ontology classes, particles, their senses and the verb arguments which the ontology does not have a way to handle currently. We leave these for future work. Finally, we demonstrated through an example how grammatical/semantic/ontological information, that enables compositional parsing, is used to obtain full semantic representation of sentences.

Acknowledgments

This work has been partially supported by the DARPA Big Mechanism program under ARO grant W911NF-14-1-0391 and the DARPA CwC program under ARO grant W911NF-15-1-0542. We would also like to thank anonymous reviewers for their useful feedback. Also we thank William de Beaumont for technical support.

References

James F. Allen and 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*.

J. Allen, M. Dzikovska, M. Manshadi, and M. Swift. 2007. Deep linguistic processing for spoken dialogue systems. In *Proceedings of the ACL 2007 Workshop on Deep Linguistic Processing*, pages 49–56. Association for Computational Linguistics.

J. Allen, M. Swift, and W. de Beaumont. 2008. Deep semantic analysis of text. In *Symposium on Semantics in Systems for Text Processing (STEP)*.

Timothy Baldwin, Colin Bannard, Takaaki Tanaka, and 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, pages 89–96, Stroudsburg, PA, USA. Association for Computational Linguistics.

Colin Bannard, Timothy Baldwin, and 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, pages 65–72, Stroudsburg, PA, USA. Association for Computational Linguistics.

Colin Bannard. 2002. Statistical techniques for automatically inferring the semantics of verb-particle constructions. Technical report.

M. Bréal. 1900. *Semantics*. Translated by Mrs. H. Cust. Henry Holt, New York.

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

E. V. Clark. 1987. The principle of contrast. In B. MacWhinney, editor, *Mechanisms of Language Acquisition*, pages 1–33. Academic Press.

Paul Cook and Suzanne Stevenson. 2006. Classifying particle semantics in english verb-particle constructions. In *Proceedings of the Workshop on Multiword Expressions: Identifying and Exploiting Underlying Properties*, pages 45–53.

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

Christiane Fellbaum. 1998. *WordNet: An Electronic Lexical Database*. MIT Press.

B. Fraser. 1976. *The Verb-Particle Combination in English*. Academic Press.

John A. Hawkins. 2000. The relative order of prepositional phrases in english: Going beyond manner-place-time. *Language Variation and Change*, 11:231–266.

Ray Jackendoff. 1997. *The Architecture of the Language Faculty*. MIT Press.

Ray Jackendoff. 2002. English particle constructions, the lexicon, and the autonomy of syntax. In *Verb-Particle Explorations*. Mouton de Gruyter.

C. Johnson and C. 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*, pages 56–62. Morgan Kaufmann Publishers Inc.

P. Kiparsky. 1983. Word-formation and the lexicon. In F. Ingemman, editor, *Proceedings of the 1982 Mid-America Linguistics Conference*, pages 47–78. University of Kansas, Dept. of Linguistics.

Karin Kipper-Schuler. 2005. *VerbNet: A broad-coverage, comprehensive verb lexicon*. Ph.D. thesis, Computer and Information Science Dept., University of Pennsylvania, Philadelphia, PA, 6.

Barbara Lohse, John A. Hawkins, and Thomas Wasow. 2004. Domain minimization in english verb-particle constructions. *Language*, 80:238–261.

Diana McCarthy, Bill Keller, and 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, pages 73–80, Stroudsburg, PA, USA. Association for Computational Linguistics.

George A. Miller. 1995. Wordnet: A lexical database for english. 38(11):39–41.

Antonio Moreno-Ortiz, Chantal Pérez-Hernández, and 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 2013)*, pages 1–10. Association for Computational Linguistics.

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

J. Patrick and J. Fletcher. 2004. Differentiating types of verb particle constructions. In *Proc. of Australasian Language Technology Workshop 2004 (ALTW2004)*.

H. Paul. 1890. *Principles of the History of Language*. Translated by H. A. Strong from the 2nd German Edition. McGrath, Reprinted College Park, MD. 1970.

Andrea Tyler and Vyvyan Evans. 2003. *The Semantics of English Prepositions: Spatial Scenes, Embodied Meaning, and Cognition*. Cambridge University Press, New York.

Aline Villavicencio. 2003. Verb-particle constructions and lexical resources. In *Proceedings of the ACL 2003 Workshop on Multiword Expressions: Analysis, Acquisition and Treatment - Volume 18*, MWE '03, pages 57–64, Stroudsburg, PA, USA. Association for Computational Linguistics.

Aline Villavicencio. 2006. Verb-particle constructions in the world wide web. In P. Saint-Dizier, editor, *Computational Linguistics Dimensions of the Syntax and Semantics of Prepositions*, pages 115–130. Springer.

Veronika Vincze, Nagy T. István, and Gábor Berend. 2011. Multiword expressions and named entities in the wiki50 corpus. In *Proceedings of Recent Advances in Natural Language Processing*, pages 289–295.

P. Vossen. 1997. Eurowordnet: a multilingual database for information retrieval. In *Proceedings of the DELOS workshop on Cross-language Information Retrieval*.