Automatic Identification of English Verb Particle Constructions using Linguistic Features

Su Nam Kim and **Timothy Baldwin**

Department of Computer Science and Software Engineering
University of Melbourne, Victoria 3010 Australia
{snkim,tim}@csse.unimelb.edu.au

Abstract

This paper presents a method for identifying token instances of verb particle constructions (VPCs) automatically, based on the output of the RASP parser. The proposed method pools together instances of VPCs and verb-PPs from the parser output and uses the sentential context of each such instance to differentiate VPCs from verb-PPs. We show our technique to perform at an F-score of 97.4% at identifying VPCs in Wall Street Journal and Brown Corpus data taken from the Penn Treebank.

1 Introduction

Multiword expressions (hereafter **MWEs**) are lexical items that can be decomposed into multiple simplex words and display lexical, syntactic and/or semantic idiosyncracies (Sag et al., 2002; Calzolari et al., 2002). In the case of English, MWEs are conventionally categorised syntactico-semantically into classes such as compound nominals (e.g. *New York*, *apple juice*, *GM car*), verb particle constructions (e.g. *hand in*, *battle on*), non-decomposable idioms (e.g. *a piece of cake*, *kick the bucket*) and light-verb constructions (e.g. *make a mistake*). MWE research has focussed largely on their implications in language understanding, fluency and robustness (Pearce, 2001; Sag et al., 2002; Copestake and Lascarides, 1997; Bannard et al., 2003; McCarthy et al., 2003; Widdows and Dorow, 2005). In this paper, our goal is to identify individual token instances of English verb particle constructions (**VPCs** hereafter) in running text.

For the purposes of this paper, we follow Baldwin (2005) in adopting the simplifying assumption that VPCs: (a) consist of a head verb and a unique prepositional particle (e.g. *hand in*, *walk off*); and (b) are either transitive (e.g. *hand in, put on*) or intransitive (e.g. *battle on*). A defining characteristic of transitive VPCs is that they can gen-erally occur with either joined (e.g. *He put on the sweater*) or split (e.g. *He put the sweater on*) word order. In the case that the object is pronominal, however, the VPC must occur in split word order (c.f. **He handed in it*) (Huddleston and Pullum, 2002; Villavicencio, 2003).

The semantics of the VPC can either derive transparently from the semantics of the head verb and particle (e.g. *walk off*) or be significantly removed from the semantics of the head verb and/or particle (e.g. *look up*); analogously, the selectional preferences of VPCs can mirror those of their head verbs or alternatively diverge markedly. The syntax of the VPC can also coincide with that of the head verb (e.g. *walk off*) or alternatively diverge (e.g. *lift off*).

In the following, we review relevant past research on VPCs, focusing on the extraction/identification of VPCs and the prediction of the compositionality/productivity of VPCs.

There is a modest body of research on the identification and extraction of VPCs. Note that in the case of VPC **identification** we seek to detect individual VPC **token** instances in corpus data, whereas in the case of VPC **extraction** we seek to arrive at an inventory of VPC **types**/lexical items based on analysis of token instances in corpus data. Li et al. (2003) identify English VPCs (or "phrasal verbs" in their parlance) using hand-coded regular expressions. Baldwin and Villavicencio (2002) extract a simple list of VPCs from corpus data, while Baldwin (2005) extracts VPCs with valence information under the umbrella of deep lexical acquisition.[1] The method of Baldwin (2005) is aimed at VPC extraction and takes into account only the syntactic features of verbs. In this paper, our interest is in VPC identification, and we make use of deeper semantic information.

In Fraser (1976) and Villavicencio (2006) it is argued that the semantic properties of verbs can determine the likelihood of their occurrence with

---

[1]The learning of lexical items in a form that can be fed directly into a deep grammar or other richly-annotated lexical resource

particles. Bannard et al. (2003) and McCarthy et al. (2003) investigate methods for estimating the compositionality of VPCs based largely on distributional similarity of the head verb and VPC. O'Hara and Wiebe (2003) propose a method for disambiguating the verb sense of verb-PPs. While our interest is in VPC identification—a fundamentally syntactic task—we draw on the shallow semantic processing employed in these methods in modelling the semantics of VPCs relative to their base verbs.

The contribution of this paper is to combine syntactic and semantic features in the task of VPC identification. The basic intuition behind the proposed method is that the selectional preferences of VPCs over predefined argument positions,[2] should provide insight into whether a verb and preposition in a given sentential context combine to form a VPC (e.g. *Kim handed in the paper*) or alternatively constitute a verb-PP (e.g. *Kim walked in the room*). That is, we seek to identify individual preposition token instances as intransitive prepositions (i.e. prepositional particles) or transitive particles based on analysis of the governing verb.

The remainder of the paper is structured as follows. Section 2 outlines the linguistic features of verbs and their co-occuring nouns. Section 3 provides a detailed description of our technique. Section 4 describes the data properties and the identification method. Section 5 contains detailed evaluation of the proposed method. Section 6 discusses the effectiveness of our approach. Finally, Section 7 summarizes the paper and outlines future work.

2 Linguistic Features

When verbs co-occur with particles to form VPCs, their meaning can be significantly different from the semantics of the head verb in isolation. According to Baldwin et al. (2003), divergences in VPC and head verb semantics are often reflected in differing selectional preferences, as manifested in patterns of noun co-occurrence. In one example cited in the paper, the cosine similarity between *cut* and *cut out*, based on word co-occurrence vectors, was found to be greater than that between *cut* and *cut off*, mirroring the intuitive compositionality of these VPCs.

(1) and (2) illustrate the difference in the selectional preferences of the verb *put* in isolation as compared with the VPC *put on*.[3]

---

[2]Focusing exclusively on the subject and object argument positions.

[3]All sense definitions are derived from WordNet 2.1.

(1) *put* = place

  EX: *Put the* $\boxed{book}$ *on the table.*
  ARGS: $book_{\text{OBJ}}$ = book, publication, object
  ANALYSIS: verb-PP

(2) *put on* = wear
  EX: *Put on the* $\boxed{sweater}$.
  ARGS: $sweater_{\text{OBJ}}$ = garment, clothing
  ANALYSIS: verb particle construction

While *put on* is generally used in the context of wearing something, it usually occurs with clothing-type nouns such as *sweater* and *coat*, whereas the simplex *put* has less sharply defined selectional restrictions and can occur with any noun. In terms of the word senses of the head nouns of the object NPs, the VPC *put on* will tend to co-occur with objects which have the semantics of clothes or garment. On the other hand, the simplex verb *put* in isolation tends to be used with objects with the semantics of *object* and prepositional phrases containing NPs with the semantics of *place*.

Also, as observed above, the valence of a VPC can differ from that of the head verb. (3) and (4) illustrate two different senses of *take off* with intransitive and transitive syntax, respectively. Note that *take* cannot occur as a simplex intransitive verb.

(3) *take off* = lift off

  EX: *The* $\boxed{airplane}$ *takes off.*
  ARGS: $airplane_{\text{SUBJ}}$ = airplane, aeroplane
  ANALYSIS: verb particle construction

(4) *take off* = remove

  EX: $\boxed{They}$ *take off the* $\boxed{cape}$.
  ARGS: $they_{\text{SUBJ}}$ = person, individual
    $cape_{\text{OBJ}}$ = garment, clothing
  ANALYSIS: verb particle construction

Note that in (3), *take off* = lift off co-occurs with a subject of the class airplane, aeroplane. In (4), on the other hand, *take off* = remove and the corresponding object noun is of class garment or clothing. From the above, we can see that head nouns in the subject and object argument positions can be used to distinguish VPCs from simplex verbs with prepositional phrases (i.e. verb-PPs).

3 Approach

Our goal is to distinguish VPCs from verb-PPs in corpus data, i.e. to take individual inputs such as *Kim handed the paper in today* and tag each as either a VPC or a verb-PP. Our basic approach is to parse each sentence with RASP (Briscoe and Carroll, 2002) to obtain a first-gloss estimate of the VPC and verb-PP token instances, and also identify the head nouns of the arguments of each VPC and simplex verb. For the head noun of each subject and object, as identified by RASP, we use WordNet 2.1 (Fellbaum, 1998) to obtain the word sense. Finally we build a supervised classifier using TiMBL 5.1 (Daelemans et al., 2004).

3.1 Method

Compared to the method proposed by Baldwin (2005), our approach (a) tackles the task of VPC identification rather than VPC extraction, and (b) uses both syntactic and semantic features, employing the WordNet 2.1 senses of the subject and/or object(s) of the verb. In the sentence *He put the coat on the table*, e.g., to distinguish the VPC *put on* from the verb *put* occurring with the prepositional phrase *on the table*, we identify the senses of the head nouns of the subject and object(s) of the verb *put* (i.e. *he* and *coat*, respectively).

First, we parse all sentences in the given corpus using RASP, and identify verbs and prepositions in the RASP output. This is a simple process of checking the POS tags in the most-probable parse, and for both particles (tagged RP) and transitive prepositions (tagged II) reading off the governing verb from the dependency tuple output (see Section 3.2 for details). We also retrieved the head nouns of the subject and object(s) of each head verb directly from the dependency tuples. Using WordNet 2.1, we then obtain the word sense of the head nouns.

The VPCs or verb-PPs are represented with corresponding information as given below:

$$P(type|v, p, ws_{\text{SUBJ}}, ws_{\text{DOBJ}}, ws_{\text{IOBJ}})$$

where *type* denotes either a VPC or verb-PP, v is the head verb, p is the preposition, and ws_* is the word sense of the subject, direct object or indirect object.

Once all the data was gathered, we separated it into test and training data. We then used TiMBL 5.1 to learn a classifier from the training data, which was then run and evaluated over the test data. See Section 5 for full details of the results.

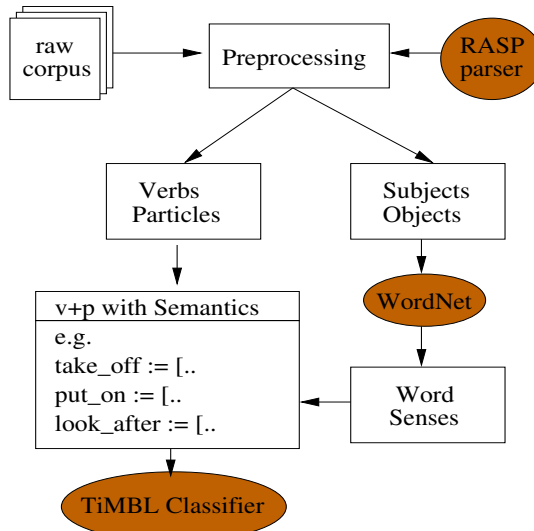Figure 1 depicts the complete process used to distinguish VPCs from verb-PPs.

Figure 1: System Architecture

3.2 On the use of RASP, WordNet and TiMBL

RASP is used to identify the syntactic structure of each sentence, including the head nouns of arguments and first-gloss determination of whether a given preposition is incorporated in a VPC or verb-PP. The RASP output contains dependency tuples derived from the most probable parse, each of which includes a label identifying the nature of the dependency (e.g. SUBJ, DOBJ), the head word of the modifying constituent, and the head of the modified constituent. In addition, each word is tagged with a POS tag from which it is possible to determine the valence of any prepositions. McCarthy et al. (2003) evaluate the precision of RASP at identifying VPCs to be 87.6% and the recall to be 49.4%. However the paper does not evaluate the parser's ability to distinguish sentences containing VPCs and sentences with verb-PPs.

To better understand the baseline performance of RASP, we counted the number of false-positive examples tagged with RP and false-negative examples tagged with II, relative to gold-standard data. See Section 5 for details.

We use WordNet to obtain the first-sense word sense of the head nouns of subject and object phrases, according to the default word sense ranking provided within WordNet. McCarthy et al. (2004) found that 54% of word tokens are used with their first (or default) sense. With the performance of current word sense disambiguation (WSD) systems hovering around 60-70%, a simple first-sense WSD system has room for improvement, but is sufficient for our immediate purposes

in this paper.

To evaluate our approach, we built a supervised classifier using the TiMBL 5.1 memory-based learner and training data extracted from the Brown and WSJ corpora.

4 Data Collection

We evaluated out method by running RASP over Brown Corpus and Wall Street Journal, as contained in the Penn Treebank (Marcus et al., 1993).

4.1 Data Classification

The data we consider is sentences containing prepositions tagged as either `RP` or `II`. Based on the output of RASP, we divide the data into four groups:
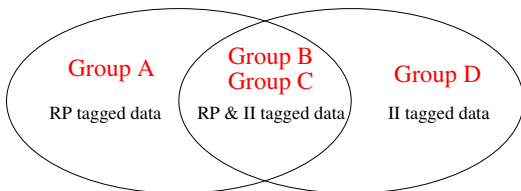
Group A contains the verb–preposition token instances tagged tagged exclusively as VPCs (i.e. the preposition is never tagged as `II` in combination with the given head verb). Group B contains the verb–preposition token instances identified as VPCs by RASP where there were also instances of that same combination identified as verb-PPs. Group C contains the verb–preposition token instances identified as verb-PPs by RASP where there were also instances of that same combination identified as VPCs. Finally, group D contains the verb-preposition combinations which were tagged exclusively as verb-PPs by RASP. We focus particularly on disambiguating verb–preposition token instances falling into groups B and C, where RASP has identified an ambiguity for that particular combination. We do not further classify token instances in group D, on the grounds that (a) for high-frequency verb–preposition combinations, RASP was unable to find a single instance warranting a VPC analysis, suggesting it had high confidence in its ability to correctly identify instances of this lexical type, and (b) for low-frequency verb–preposition combinations where the confidence of there definitely no being a VPC usage is low, the token sample is too small to disambiguate effectively and the overall impact would be negligible even if we tried. We do, however, return to considered data in group D in computing the precision and recall of RASP.

Naturally, the output of RASP parser is not error-free, i.e. VPCs may be parsed as verb-PPs

	FPR	FNR	Agreement
Group A	4.08%	—	95.24%
Group B	3.96%	—	99.61%
Group C	—	10.15%	93.27%
Group D	—	3.4%	99.20%

Table 1: False positive rate (FPR), false negative rate (FNR) and inter-annotator agreement across the four groups of token instances

	$f \geq 1$		$f \geq 5$	
	VPC	V-PP	VPC	V-PP
Group A	5,223	0	3,787	0
Group B	1,312	0	1,108	0
Group C	0	995	0	217
Total	6,535	995	4,895	217

Table 2: The number of VPC and verb-PP token instances occurring in groups A, B and C at varying frequency cut-offs

and vice versa. In particular, other than the reported results of McCarthy et al. (2003) targeting VPCs vs. all other analyses, we had no a priori sense of RASP's ability to distinguish VPCs and verb-PPs. Therefore, we manually checked the false-positive and false-negative rates in all four groups and obtained the performance of parser with respect to VPCs. The verb-PPs in group A and B are false-positives while the VPCs in group C and D are false-negatives (we consider the VPCs to be positive examples).

To calculate the number of incorrect examples, two human annotators independently checked each verb–preposition instance. Table 1 details the rate of false-positives and false-negative examples in each data group, as well as the inter-annotator agreement (calculated over the entire group).

4.2 Collection

We combined together the 6,535 (putative) VPCs and 995 (putative) verb-PPs from groups A, B and C, as identified by RASP over the corpus data. Table 2 shows the number of VPCs in groups A and B and the number of verb-PPs in group C. The first number is the number of examples occuring at least once and the second number that of examples occurring five or more times.

From the sentences containing VPCs and verb-PPs, we retrieved a total of 8,165 nouns, including

Type	Groups A&B	Group C
common noun	7,116	1,239
personal pronoun	629	79
demonstrative pronoun	127	1
proper noun	156	18
who	94	6
which	32	0
No sense (*what*)	11	0

Table 3: Breakdown of subject and object head nouns in group A&B, and group C

pronouns (e.g. *I, he, she*), proper nouns (e.g. *CITI, Canada, Ford*) and demonstrative pronouns (e.g. *one, some, this*), which occurred as the head noun of a subject or object of a VPC in group A or B. We similarly retrieved 1,343 nouns for verb-PPs in group C. Table 3 shows the distribution of different noun types in these two sets.

We found that about 10% of the nouns are pronouns (personal or demonstrative), proper nouns or WH words. For pronouns, we manually resolved the antecedent and took this as the head noun. When *which* is used as a relative pronoun, we identified if it was coindexed with an argument position of a VPC or verb-PP, and if so, manually identified the antecedent, as illustrated in (5).

(5) **EX:** *Tom likes the* $\boxed{books}$ *which* $\boxed{he}$ $\underline{sold}$ *off*.

 ARGS: he_{SUBJ} = person

 $which_{\text{OBJ}}$ = book

With *what*, on the other hand, we were generally not able to identify an antecedent, in which case the argument position was left without a word sense (we come back to this in Section 6).

(6) *Tom didn't* $\underline{look}$ *up* $\boxed{what}$ *to do.*
 $\boxed{What}$ $\underline{went}$ $\underline{on}$?

We also replaced all proper nouns with corresponding common noun hypernyms based on manual disambiguation, as the coverage of proper nouns in WordNet is (intentionally) poor. The following are examples of proper nouns and their common noun hypernyms:

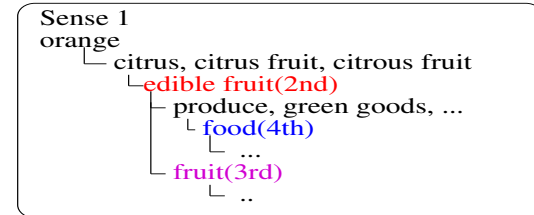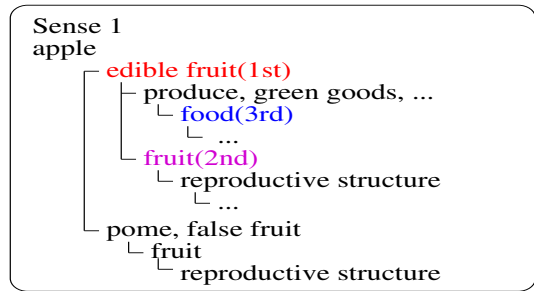| Proper noun	Common noun hypernym
CITI	bank
Canada	country
Ford	company
Smith	human

Figure 2: Senses of *apple* and *orange*

When we retrieved the first word sense of nouns from WordNet, we selected the first sense and the associated hypernyms (up to) three levels up the WordNet hierarchy. This is intended as a crude form of smoothing for closely-related word senses which occur in the same basic region of the WordNet hierarchy. As an illustration of this process, in Figure 2, *apple* and *orange* are used as edible fruit, fruit or food, and the semantic overlap is picked up on by the fact that edible fruit is a hypernym of both *apple* and *orange*. On the other hand, food is the fourth hypernym for *orange* so it is ignored by our method. However, because we use the four senses, the common senses of nouns are extracted properly. This approach works reasonably well for retrieving common word senses of nouns which are in the immediate vicinity of each other in the WordNet hierarchy, as was the case with *apple* and *orange*. In terms of feature representation, we generate an individual instance for each noun sense generated based on the above method, and in the case that we have multiple arguments for a given VPC or verb-PP (e.g. both a subject and a direct object), we generate an individual instance for the cross product of all sense combinations between the arguments.

We use 80% of the data for training and 20% for testing. The following is the total number of training instances, before and after performing hypernym expansion:

	Training Instances	
	Before expansion	After expansion
Group A	5,223	24,602
Group B	1,312	4,158
Group C	995	5,985

Group	Frequency of VPCs	Size
B	$(f_{\geq 1})$	test:272
	$(f_{\geq 5})$	train:1,040
BA	$(f_{\geq 1} \ \& \ f_{\geq 1})$	test:1,327
	$(f_{\geq 5} \ \& \ f_{\geq 5})$	train:4,163
BC	$(f_{\geq 1} \ \& \ f_{\geq 1})$	test:498
	$(f_{\geq 5} \ \& \ f_{\geq 1})$	train:1,809
BAC	$(f_{\geq 1} \ \& \ f_{\geq 1} \ \& \ f_{\geq 1})$	test:1,598
	$(f_{\geq 5} \ \& \ f_{\geq 5} \ \& \ f_{\geq 1})$	train:5,932

Table 4: Data set sizes at different frequency cut-offs

Data	Freq	P	R	F
RASP	$f_{\geq 1}$	.959	.955	.957
B	$f_{\geq 1}$	1.0	.819	.901
	$f_{\geq 5}$	1.0	.919	.957
BA	$f_{\geq 1} \ f_{\geq 1}$	1.0	.959	.979
	$f_{\geq 5} \ f_{\geq 5}$	1.0	.962	.980
BC	$f_{\geq 1} \ f_{\geq 1}$	.809	.845	.827
	$f_{\geq 5} \ f_{\geq 1}$	.836	.922	.877
BAC	$f_{\geq 1} \ f_{\geq 1} \ f_{\geq 1}$	.962	.962	.962
	$f_{\geq 5} \ f_{\geq 5} \ f_{\geq 1}$	.964	.983	.974

Table 5: Results for VPC identification only (P = precision, R = recall, F = F-score)

Data	Freq	Type	P	R	F
RASP	$f_{\geq 1}$	P+V	.933	–	–
BC	$f_{\geq 1} \ f_{\geq 1}$	P+V	.8068	.8033	.8051
	$f_{\geq 5} \ f_{\geq 1}$	P+V	.8653	.8529	.8591
BAC	$f_{\geq 1} \ f_{\geq 1}$	P+V	.8660	.8660	.8660
	$f_{\geq 5} \ f_{\geq 1}$	P+V	.9272	.8836	.9054

Table 6: Results for VPC (=V) and verb-PP (=P) identification (P = precision, R = recall, F = F-score)

5 Evaluation

We selected 20% of the test data from different combinations of the four groups and over the two frequency thresholds, leading to a total of 8 test data sets. The first data set contains examples from group B only, the second set is from groups B and A, the third set is from groups B and C, and the fourth set is from groups B, A and C. Additionally, each data set is divided into: (1) $f \geq 1$, i.e. verb–preposition combinations occurring at least once, and (2) $f \geq 5$, i.e. verb–preposition combinations occurring at least five times (hereafter, $f \geq 1$ is labelled $f_{\geq 1}$ and $f \geq 5$ is labelled $f_{\geq 5}$). In the group C data, there are 217 verb-PPs with $f_{\geq 5}$, which is slightly more than 20% of the data so we use verb-PPs with $f_{\geq 1}$ for experiments instead of verb-PP with $f_{\geq 5}$. The first and second data sets do not contain negative examples while the third and fourth data sets contain both positive and negative examples. As a result, the precision for the first two data sets is 1.0.

Table 5 shows the precision, recall and F-score of our method over each data set, relative to the identification of VPCs only. *A,B,C* are groups and *f#* is the frequency of examples.

Table 6 compares the performance of VPC identification and verb-PP identification.

Table 7 indicates the result using four word senses (i.e. with hypernym expansion) and only one word sense (i.e. the first sense only).

6 Discussion

The performance of RASP as shown in Tables 5 and 6 is based on human judgement. Note that we only consider the ability of the parser to distinguish sentences with prepositions as either VPCs or verb-PPs (i.e. we judge the parse to be correct if the preposition is classified correctly, irrespective of whether there are other errors in the output).

Also, we ignore the ambiguity between particles and adverbs, which is the principal reason for our evaluation being much higher than that reported by McCarthy et al. (2003). In Table 5, the precision (P) and recall (R) for VPCs are computed as follows:

$$P = \frac{\text{Data Correctly Tagged as VPCs}}{\text{Data Retrieved as VPCs}}$$

$$R = \frac{\text{Data Correctly Tagged as VPCs}}{\text{All VPCs in Data Set}}$$

The performance of RASP in Table 6 shows how well it distinguishes between VPCs and verb-PPs for ambiguous verb–preposition combinations. Since Table 6 shows the comparative performance of our method between VPCs and verb-PPs, the performance of RASP with examples which are misrecognized as each other should be the guideline. Note, the baseline RASP accuracy, based on assigning the majority class to instances in each of groups A, B and C, is 83.04%.

In Table 5, the performance over high-frequency data identified from groups B, A and C is the highest (F-score = .974). In general, we would expect the data set containing the high frequency and both positive and negative examples

Freq	Type	#	P	R	F
$f_{\geq 1}$	V	4WS	.962	.962	.962
		1WS	.958	.969	.963
$f_{\geq 1}$	P	4WS	.769	.769	.769
		1WS	.800	*.743*	.770
$f_{\geq 5}$	V	4WS	.964	.983	.974
		1WS	.950	.973	.962
$f_{\geq 5}$	P	4WS	.889	.783	.832
		1WS	.813	.614	.749

Table 7: Results with hypernym expansion (4WS) and only the first sense (1WS), in terms of precision (P), recall (R) and F-score (F)

Figure 3: Error rate reduction for VPCs of varying compositionality

to give us the best performance at VPC identification. We achieved a slightly better result than the 95.8%-97.5% performance reported by Li et al. (2003). However, considering that Li et al. (2003) need considerable time and human labour to generate hand-coded rules, our method has advantages in terms of both raw performance and labour efficiency.

Combining the results for Table 5 and Table 6, we see that our method performs better for VPC identification than verb-PP identification. Since we do not take into account the data from group D with our method, the performance of verb-PP identification is low compared to that for RASP, which in turn leads to a decrement in the overall performance.

Since we ignored the data from group D containing unambiguous verb-PPs, the number of positive training instances for verb-PP identification was relatively small. As for the different number of word senses in Table 7, we conclude that the more word senses the better the performance, particularly for higher-frequency data items.

In order to get a clearer sense of the impact of selectional preferences on the results, we investigated the relative performance over VPCs of varying semantic compositionality, based on 117 VPCs ($f_{\geq 1}$) attested in the data set of McCarthy et al. (2003). According to our hypothesis from above, we would expect VPCs with low compositionality to have markedly different selectional preferences to the corresponding simplex verb, and VPCs with high compositionality to have similar selectional preferences to the simplex verb. In terms of the performance of our method, therefore, we would expect the degree of compositionality to be inversely proportional to the system performance. We test this hypothesis in Figure 3, where we calculate the error rate reduction (in F-score)
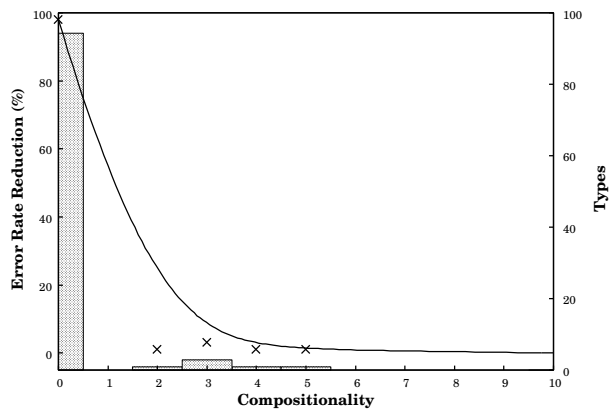
for the proposed method relative to the majority-class baseline, at various degrees of compositionality. McCarthy et al. (2003) provides compositionality judgements from three human judges, which we take the average of and bin into 11 categories (with 0 = non-compositional and 10 = fully compositional). In Figure 3, we plot both the error rate reduction in each bin (both the raw numbers and a smoothed curve), and also the number of attested VPC types found in each bin. From the graph, we see our hypothesis born out that, with perfect performance over non-compositional VPCs and near-baseline performance over fully compositional VPCs. Combining this result with the overall results from above, we conclude that our method is highly successful at distinguishing non-compositional VPCs from verb-PPs, and further that there is a direct correlation between the degree of compositionality and the similarity of the selectional preferences of VPCs and their verb counterparts.

Several factors are considered to have influenced performance. Some data instances are missing head nouns which would assist us in determining the semantics of the verb–preposition combination. Particular examples of this are imperative and abbreviated sentences:

(7) a. *Come in.*
     b. *(How is your cold?) Broiled out.*

Another confounding factor is the lack of word sense data, particularly in WH questions:

(8) a. What *do I hand in?*
     b. *You can add up* anything .

71

7 Conclusion

In this paper, we have proposed a method for identifying VPCs automatically from raw corpus data. We first used the RASP parser to identify VPC and verb-PP candidates. Then, we used analysis of the head nouns of the arguments of the head verbs to model selectional preferences, and in doing so, distinguish between VPCs and verb-PPs. Using TiMBL 5.1, we built a classifier which achieved an F-score of 97.4% at identifying frequent VPC examples. We also investigated the comparative performance of RASP at VPC identification.

The principal drawback of our method is that it relies on the performance of RASP and we assume a pronoun resolution oracle to access the word senses of pronouns. Since the performance of such systems is improving, however, we consider our approach to be a promising, stable method of identifying VPCs.

Acknowledgements

References

Timothy Baldwin and Aline Villavicencio. 2002. Extracting the unextractable: A case study on verb-particles. In *Proc. of the 6th Conference on Natural Language Learning (CoNLL-2002)*, pages 98–104, Taipei, Taiwan.

Timothy Baldwin, Colin Bannard, Takaaki Tanaka, and Dominic Widdows. 2003. An empirical model of multiword expression decomposability. In *Proc. of the ACL-2003 Workshop on Multiword Expressions: Analysis, Acquisition and Treatment*, pages 89–96, Sapporo, Japan.

Timothy Baldwin. 2005. The deep lexical acquisition of English verb-particle constructions. *Computer Speech and Language, Special Issue on Multiword Expressions*, 19(4):398–414.

Colin Bannard, Timothy Baldwin, and Alex Lascarides. 2003. A statistical approach to the semantics of verb-particles. In *Proc. of the ACL-2003 Workshop on Multiword Expressions: Analysis, Acquisition and Treatment*, pages 65–72, Sapporo, Japan.

Ted Briscoe and John Carroll. 2002. Robust accurate statistical annotation of general text. In *Proc. of the 3rd International Conference on Language Resources and Evaluation (LREC 2002)*, pages 1499–1504, Las Palmas, Canary Islands.

Nicoletta Calzolari, Charles Fillmore, Ralph Grishman, Nancy Ide, Alessandro Lenci, Catherine MacLeod, and Antonio Zampolli. 2002. Towards best practice for multiword expressions in computational lexicons. In *Proc. of the 3rd International Conference on Language Resources and Evaluation (LREC 2002)*, pages 1934–40, Las Palmas, Canary Islands.

Ann Copestake and Alex Lascarides. 1997. Integrating symbolic and statistical representations: The lexicon pragmatics interface. In *Proc. of the 35th Annual Meeting of the ACL and 8th Conference of the EACL (ACL-EACL'97)*, pages 136–43, Madrid, Spain.

Walter Daelemans, Jakub Zavrel, Ko van der Sloot, and Antal van den Bosch. 2004. *TiMBL: Tilburg Memory Based Learner, version 5.1, Reference Guide*. ILK Technical Report 04-02.

Christiane Fellbaum, editor. 1998. *WordNet: An Electronic Lexical Database*. MIT Press, Cambridge, USA.

B. Fraser. 1976. *The Verb-Particle Combination in English*. The Hague: Mouton.

Rodney Huddleston and Geoffrey K. Pullum. 2002. *The Cambridge Grammar of the English Language*. Cambridge University Press, Cambridge, UK.

Wei Li, Xiuhong Zhang, Cheng Niu, Yuankai Jiang, and Rohini K. Srihari. 2003. An expert lexicon approach to identifying English phrasal verbs. In *Proc. of the 41st Annual Meeting of the ACL*, pages 513–20, Sapporo, Japan.

Mitchell P. Marcus, Beatrice Santorini, and Mary Ann Marcinkiewicz. 1993. Building a large annotated corpus of English: the Penn treebank. *Computational Linguistics*, 19(2):313–30.

Diana McCarthy, Bill Keller, and John Carroll. 2003. Detecting a continuum of compositionality in phrasal verbs. In *Proc. of the ACL-2003 Workshop on Multiword Expressions: Analysis, Acquisition and Treatment*, Sapporo, Japan.

Diana McCarthy, Rob Koeling, Julie Weeds, and John Carroll. 2004. Finding predominant senses in untagged text. In *Proc. of the 42nd Annual Meeting of the ACL*, pages 280–7, Barcelona, Spain.

Tom O'Hara and Janyce Wiebe. 2003. Preposition semantic classification via Treebank and FrameNet. In *Proc. of the 7th Conference on Natural Language Learning (CoNLL-2003)*, pages 79–86, Edmonton, Canada.

Darren Pearce. 2001. Synonymy in collocation extraction. In *Proceedings of the NAACL 2001 Workshop on WordNet and Other Lexical Resources: Applications, Extensions and Customizations*, Pittsburgh, USA.

Ivan A. Sag, Timothy Baldwin, Francis Bond, Ann Copestake, and Dan Flickinger. 2002. Multiword expressions: A pain in the neck for NLP. In *Proc. of the 3rd International Conference on Intelligent Text Processing and Computational Linguistics (CICLing-2002)*, pages 1–15, Mexico City, Mexico.

Aline Villavicencio. 2003. Verb-particle constructions and lexical resources. In *Proc. of the ACL-2003 Workshop on Multiword Expressions: Analysis, Acquisition and Treatment*, pages 57–64, Sapporo, Japan.

Aline Villavicencio. 2006. Verb-particle constructions in the world wide web. In Patrick Saint-Dizier, editor, *Computational Linguistics Dimensions of Syntax and Semantics of Prepositions*. Springer, Dordrecht, Netherlands.

Dominic Widdows and Beate Dorow. 2005. Automatic extraction of idioms using graph analysis and asymmetric lexicosyntactic patterns. In *Proc. of the ACL-SIGLEX 2005 Workshop on Deep Lexical Acquisition*, pages 48–56, Ann Arbor, USA.