Automatically Classifying English Verb-Particle Constructions by Particle Semantics

by

Christopher Paul Cook

A thesis submitted in conformity with the requirements
for the degree of Master of Science
Graduate Department of Computer Science
University of Toronto

Abstract

Automatically Classifying English Verb-Particle Constructions by Particle Semantics

Christopher Paul Cook

Master of Science

Graduate Department of Computer Science

University of Toronto

2006

We address the issue of automatically determining the semantic contribution of the particle in a verb-particle construction (VPC), a task which has been previously ignored in computational work on VPCs. Adopting a cognitive linguistic standpoint, we assume that every VPC is compositional, and that the semantic contribution of a particle corresponds to one of a small number of senses. We develop a feature space based on syntactic and semantic properties of verbs and VPCs for type classification of English VPCs according to the sense contributed by their particle. We focus on VPCs using the particle up since it is very frequent and exhibits a wide range of meanings. In our experiments on unseen test VPCs, features which are motivated by properties specific to verbs and VPCs outperform linguistically uninformed word co-occurrence features, and give a reduction in error rate of around 20–30% over a chance baseline.

Dedication

To my parents for their love and support.

Acknowledgements

I must first thank my supervisor, Suzanne Stevenson, for her invaluable advice, support and encouragement. Without her this thesis would not have been possible.

I must also thank Graeme Hirst, my second reader, for his many insightful comments which greatly improved the quality of this thesis.

I am indebted to Eric Joanis for providing, and offering such tremendous support for, ExtractVerb, the software used to calculate many of the features used in this study. I also extend my gratitude to David James for his help in using his software package runLearner, the machine learning software used in this work.

I would like to thank Afra Alishahi, Afsaneh Fazly, Saif Mohammad, Robert Swier and Vivian Tsang for their many helpful discussions and comments. I would also like to thank the rest of the computational linguistics group at the University of Toronto for their support and feedback on this work.

Finally, I would like to thank the Natural Sciences and Engineering Research Council of Canada for their financial support.

Contents

1 Introduction **1**

1.1 Background 1

1.2 Statement of Purpose 2

1.3 Outline of Study 4

2 Syntax and Semantics of VPCs **6**

2.1 Syntactic Properties of VPCs 6

2.1.1 VPCs 6

2.1.2 Prepositional Verbs 9

2.1.3 Phrasal-Prepositional Verbs 10

2.2 VPC Semantics 10

2.2.1 Traditional Approaches to VPC Semantics 10

2.2.2 Cognitive Linguistic Approaches to VPC Semantics 12

3 Related Work **18**

3.1 Research on VPCs 18

3.1.1 Identification of VPCs 18

3.1.2 VPC Compositionality 19

3.1.3 VPC Productivity 23

3.2 Verb Classification 24

3.3 Preposition Semantics 25

4 Computational Models of Particle Semantics 28

4.1 Features Used in Classification . . . 28

4.1.1 Linguistic Features . . . 29

4.1.2 Word Co-occurrence Features . . . 33

4.2 The Sense Classes Used for Our Study . . . 34

5 Materials and Methods 36

5.1 Experimental Expressions . . . 36

5.2 Calculation of the Features . . . 38

5.2.1 VPC Identification . . . 38

5.2.2 Linguistic Feature Calculation . . . 38

5.2.3 Word Co-occurrence Feature Calculation . . . 43

5.2.4 Feature Contexts . . . 45

5.3 Experimental Classes . . . 45

5.4 Evaluation Metrics . . . 46

5.5 Classifier Software . . . 47

6 Experiments and Results 49

6.1 Experiments Using Linguistic Features . . . 50

6.1.1 Particle Features . . . 50

6.1.2 Slot Features . . . 51

6.1.3 Slot + Particle Features . . . 51

6.1.4 All Linguistic Features . . . 52

6.2 Experiments Using WCFs . . . 53

6.3 Experiments Combining Linguistic Features and WCFs . . . 53

6.4 Discussion of Results . . . 54

7 Conclusions 57

7.1 Summary of Contributions . . . 57

7.2 Limitations and Future Work 58

A Human Judgements of Particle Sense Contribution **61**

B Additional Experimental Results **65**

Bibliography **70**

List of Tables

5.1 Number of VPCs in frequency range of base verb. 37

5.2 High-frequency prepositions, taken from Joanis (2002), with the exception of up. 40

5.3 Prepositions for which an alternative spelling is allowed, taken from Joanis (2002). 40

5.4 Preposition groups, taken from Joanis (2002). 42

5.5 High-frequency adverbs in verb chunk. 43

5.6 High-frequency adverbs following verb chunk. 43

5.7 High-frequency particles and their frequency in the BNC. 44

5.8 Frequency of items in each sense class. 45

5.9 Frequency of items in each class for the 3-way task. 46

5.10 Frequency of items in each class for the 2-way task. 46

6.1 Accuracy (%) using linguistic features. 50

6.2 Accuracy (%) using WCFs. 53

6.3 Accuracy (%) combining slot and particle features with WCFs. 54

6.4 Accuracy (%) combining all linguistic features with WCFs. 54

A.1 Human annotator judgements for training set. 62

A.2 Human annotator judgements for verification set. 63

A.3 Human annotator judgements for test set. 64

B.1 Accuracy (%) using linguistic features. 66

B.2 Accuracy (%) using WCFs. 67

B.3 Accuracy (%) combining slot and particle features with WCFs. Note that the slot and particle features are calculated over all three contexts, while the WCFs are calculated for the specific contexts given above. 68

B.4 Accuracy (%) combining all linguistic features with WCFs. Note that the linguistic features are calculated over all three contexts, while the WCFs are calculated for the specific contexts given above. 69

List of Figures

2.1 Schema for Vert-up. 13

2.2 Schema for Goal-up. 15

2.3 Schema for Refl-up. 16

2.4 Simplified schematic network for up. 17

Chapter 1

Introduction

1.1 Background

Multiword expressions (MWEs), defined as "idiosyncratic interpretations that cross word boundaries" (Sag et al., 2002), include a wide range of phenomena such as fixed expressions (e.g., to and fro, *ad hoc*), idioms (e.g., *kick the bucket*, *spill the beans*), and light verb constructions (e.g., *take a walk*, *give a smile*). Recent work on lexical knowledge acquisition has recognized the important role of applying computational learning techniques to MWEs (Sag et al., 2002; Villavicencio et al., 2005). However, the learning of semantic properties of MWEs poses a particular challenge because of the varying degrees of their *compositionality*—the contribution of each component word to the overall semantics of the expression. MWEs fall on a continuum from fully compositional (i.e., each component contributes its meaning, as in *frying pan*) to non-compositional or idiomatic (as in *hold water*). Because of this variation, researchers have explored automatic methods for learning whether, or the degree to which, an MWE is compositional (e.g., Lin, 1999; McCarthy et al., 2003; Bannard, 2005; Fazly et al., 2005).

However, such work leaves unaddressed the basic issue of which of the possible meanings of a component word is contributed when a MWE is (at least partly) compositional.

Words are notoriously ambiguous, so that even if it can be determined that an MWE is compositional, its meaning is still unknown, since the actual semantic contribution of the components is yet to be determined. We address this problem in the domain of verb-particle constructions (VPCs) in English, a rich source of MWEs.

VPCs combine a verb with any of a finite set of particles, as in *jump up*, *figure out*, or *give in*. Particles such as *up*, *out*, or *in*, with their literal meaning based in physical spatial relations, show a variety of metaphorical and aspectual meaning extensions, as exemplified here for the particle *up*:

(1a) The sun just <u>came up</u>.[1] [vertical spatial movement]

(1b) She <u>walked up</u> to him. [movement toward a goal]

(1c) <u>Drink up</u> your juice. [completion]

(1d) He <u>curled up</u> into a ball. [reflexive movement]

Cognitive linguistic analysis, as by Langacker (1987) and Lindner (1981), can provide the basis for elaborating this type of semantic variation.

1.2 Statement of Purpose

In this study, our goal is to automatically determine the meaning of a particle when used with a given verb in a VPC. We classify English VPCs using the particle *up* according to their particle sense. In doing so, we adopt a cognitive linguistic standpoint, and assume that every VPC is (at least somewhat) compositional. We base our classification on the senses of *up* identified by Lindner's (1981) cognitive analysis of VPCs using this particle.

We hypothesize that the semantic contribution of a particle when used in a VPC with a given verb is related to that verb's meaning. We therefore develop the following

[1]In examples of VPCs, the verb and particle will be underlined.

statistical features which are motivated by specific semantic and syntactic properties of verbs. Note that we will refer to the verb in a VPC as the base verb of that VPC.

Slot Features capture the semantics of a VPC by measuring the frequency of occurrence of the base verb of the VPC with an argument or adjunct in various syntactic slots.

Adverb Features also capture the semantics of a VPC, but do so based on the frequency of co-occurrence of the base verb of the VPC and adverbs.

Nominal Features exploit the fact that the meaning of a VPC may be related to how frequently its base verb is used as a noun.

We also hypothesize that patterns of co-occurrence of the base verb of a VPC and particles are indicative of the semantics of the VPC. The following set of features captures this hypothesis.

Particle Features indicate the semantics of a VPC by measuring the ability of the base verb of the VPC to combine with other particles, and the ability of the VPC to be used with other words occurring between its verb and particle.

We contrast these features with simple word co-occurrence features, which are often used to indicate the semantics of a target word. We show that our features which are motivated by syntactic and semantic properties of verbs and particles perform best, and give a substantial reduction in error rate over a chance baseline.

In our experiments, we focus on VPCs using the particle *up*. *Up* is the most frequent particle in the corpus used in this study; therefore by focusing on *up*, we avoid problems associated with data sparseness that may plague studies of less frequent particles. *Up* also exhibits a wide range of meanings, as exemplified in sentences (1a–d), giving us the opportunity to explore classification tasks using differing numbers of classes. Although we focus on *up*, it is worth emphasizing that, for the most part, our feature space draws on general properties of VPCs, and is not specific to this particle.

A VPC may be ambiguous, with its particle occurring in more than one sense. For example, in contrast to (1a) above, *up* may contribute the goal-oriented sense to the expression *come up*, as in *The deadline is coming up.* While our long-term goal of VPC research is token classification (disambiguation) of a particle instance in context, following other recent work on VPCs (e.g., Bannard et al., 2003; McCarthy et al., 2003), we focus on the task of type classification—i.e., classification of the use of an expression, in our case a VPC with *up*, across a corpus. Given our use of features which capture the statistical behaviour relevant to a VPC across the corpus, we assume that the outcome of type classification yields the predominant sense of the particle in the VPC. Predominant sense identification is a useful component of sense disambiguation of word tokens (McCarthy et al., 2004), and we presume our VPC type classification work will contribute to later token disambiguation.

1.3 Outline of Study

This study continues as follows:

Chapter 2: Syntax and Semantics of VPCs situates VPCs with respect to other similar constructions, and discusses many of their syntactic properties. We then turn to consider the semantics of VPCs, and after a brief introduction to cognitive linguistics, give a detailed cognitive linguistic account of the semantics of VPCs using the particle *up*.

Chapter 3: Related Work examines previous work on computational approaches to verb-particle constructions. This chapter also discusses relevant pieces of work on preposition semantics and the semantic classification of verbs.

Chapter 4: Computational Models of Particle Semantics describes the two sets of features, linguistic features and word co-occurrence features, used in our experiments. The first set of features, the linguistic features, are motivated by syntactic and semantic properties of VPCs introduced in Chapter 2. This chapter concludes by describing the the sense classes of *up* used in our experiments.

Chapter 5: Materials and Methods describes the data used in our experiments, the calculation of the machine learning features using corpus statistics, and the evaluation metrics and classification software used in this study.

Chapter 6: Experiments and Results presents and discusses our results.

Chapter 7: Conclusions summarizes the contributions of this study and discusses some of its limitations.

Chapter 2

Syntax and Semantics of VPCs

In this chapter we will examine some of the syntactic properties of VPCs and two contrasting analyses of their semantics.

2.1 Syntactic Properties of VPCs

Many studies have examined English VPCs (Bolinger, 1971; Fraser, 1976; Biber et al., 1999; Dehé et al., 2002). Here we will consider the description given by Biber et al., since it situates VPCs with respect to other similar constructions.

A multiword lexical verb (MLV) is the combination of a verb with one or more words which functions similarly to a simple verb. MLVs come in several flavours of which three are relevant to this work: VPCs, prepositional verbs, and phrasal-prepositional verbs.

2.1.1 VPCs

A VPC is composed of a verb and particle.[1] Particles form a distinct syntactic category, but are generally homonymous with a subset of the prepositions; examples are *up*, *down*,

[1]Biber et al. (1999) distinguish between "phrasal verbs" and "free combinations", both of which are composed of a verb and particle. However, a phrasal verb is interpreted as a single semantic unit, while both the verb and particle in a free combination contribute their meaning. We will refer to both phrasal verbs and free combinations as VPCs, since we are interested in any combination of a verb and particle.

in, *out*, *on*, and *off*. Sentences 2 and 3 contain examples of VPCs.

(2) I <u>looked up</u> the number.

(3) She <u>turned off</u> the lights.

Since particles and prepositions are homonymous, the interpretation of a given verb and particle/preposition can be ambiguous between a VPC and a verb with a prepositional phrase adjunct or argument (V+PP). Consider the following two sentences which use the particles from sentences 2 and 3 as prepositions.

(4) I looked up the chimney.

(5) She turned off the freeway.

One property of VPCs which can be used to distinguish them syntactically from V+PPs is that a transitive VPC may appear in both the split and joined constructions. In the split construction, shown in sentences 6 and 7, one or more words occur between the verb and particle, while in the joined construction the verb and particle occur together, as in sentences 2 and 3.

(6) I <u>looked</u> the number <u>up</u>.

(7) She <u>turned</u> the lights <u>off</u>.

The split construction is not grammatical for a V+PP, as is shown by the following sentences.

(8) *I looked the chimney up.

(9) *She turned the freeway off.

The split construction is mandatory when the object of a transitive VPC is a pronoun as in:

(10) I looked it up.

(11) *I looked up it.

However, when the object of a transitive VPC is a heavy noun phrase, the joined construction is preferred:

(12) I looked up the number for the house with the white door on Bloor Street.

(13) ?I looked the number for the house with the white door on Bloor Street up.

Studies examining the split and joined constructions have determined that a number of factors relating to processing efficiency affect particle placement (Hawkins, 1994; Gries, 2002; Lohse, 2004). In some cases, a third construction in which the particle occurs before the verb, as in Sentences 14 and 15, is acceptable.

(14) Up walked the delivery man.

(15) On went the boring lecture.

This construction is discussed at length by Cappelle (2002), but will not be further considered in this study.

Another important syntactic feature of VPCs is that an adverb may be allowed between the verb and particle as in sentences 16 and 17; however, for transitive VPCs this is only acceptable in the split construction, and the adverb must appear to the right of the direct object.

(16) The handle broke *clean* off.

(17) I looked the number *right* up.

2.1.2 Prepositional Verbs

Prepositional verbs (PVs) are composed of a verb and preposition and come in two varieties:

- verb + preposition + NP
- verb + NP + preposition + NP

We will only further consider the first type of PV, since the interpretation of a verb + particle/preposition + noun-phrase may be ambiguous between a PV and a transitive VPC. Examples of prepositional verbs are:

(18) Alfred relied on his parents for financial support.

(19) The scientist referred to Wikipedia.

One property which may be used to distinguish PVs from transitive VPCs is that PVs do not allow the split construction; the verb and preposition occur joined even when the noun-phrase complement is a pronoun, as in the following sentences.

(20) Alfred relied on them.

(21) *Alfred relied them on.

(22) He referred to it.

(23) *He referred it to.

The interpretation of a verb + preposition + noun-phrase may also be ambiguous between a PV and a V+PP. These may be distinguished since PVs are more acceptable when used in the passive form, as shown below.

(24) His parents were relied on

(25) Wikipedia was referred to.

2.1.3 Phrasal-Prepositional Verbs

Phrasal-prepositional verbs (PPVs) are typically composed of a verb, particle and preposition, and require a noun-phrase direct object. Examples of PPVs are given in sentences 26 and 27.

(26) Mary came up with an idea.

(27) John stood up to his father.

PPVs are similar in structure to intransitive VPCs which take a prepositional phrase argument. This analysis of PPV structure accounts for the lack of particle movement which PPVs exhibit.

In this work we are concerned with constructions involving a verb and particle. We will use the term VPC to refer to both VPCs and PPVs, and will not distinguish between these two types of MLVs.

2.2 VPC Semantics

In the following subsections we will examine two analyses of VPC semantics: a traditional linguistic analysis and a cognitive linguistic analysis.

2.2.1 Traditional Approaches to VPC Semantics

Accounts of the semantics of VPCs are given by Bolinger (1971), Fraser (1976), Jackendoff (2002) and others. Here we choose to focus on the analysis of Jackendoff.

Jackendoff gives a three-way classification of VPCs according to the semantics contributed by their particle. The three classes which Jackendoff identifies are directional, aspectual and idiomatic.[2] In directional VPCs, examples of which are given below, the particle contributes its basic directional sense.

(28) Work just keeps piling up.

(29) The artist took the painting down.

The following sentences contain examples of aspectual VPCs.

(30) The World's oil supplies will be used up by 2050.

(31) The carpenter banged away at the nail.

In sentence (30) the particle contributes a notion of completeness, while in sentence (31) the particle contributes the notion that the action is continuing—in this case in an iterative manner. The distinguishing property of idiomatic VPCs is that the meaning of the particle cannot be easily identified. According to Jackendoff, examples of idiomatic VPCs are:

(32) The clown blew up the balloon.

(33) Fred saw a spider and freaked out.

One problem with Jackendoff's analysis of VPC semantics is that it classifies many VPCs as idiomatic; however, some of the particles in these VPCs seem to contribute, to varying degrees, a meaning. For example, in sentence (32) there is a sense in which the balloon is expanding, and therefore may be interpreted as being or becoming *up*. In sentence (33) there is a sense in which Fred is exiting from his normal mental state, and therefore may be *out*. Such shortcomings in this analysis motivate further investigation of VPC semantics. McIntyre (2002) uses the idea of construction-specific senses of particles to account for the meaning of many VPCs, but still considers some VPCs to be idiomatic.

[2] Jackendoff also identifies the *time away* construction. This construction is composed of a verb, a time expression and *away*, as in *Mary drank the night away* and *John slept the whole day away*. This construction will not be analyzed here.

2.2.2 Cognitive Linguistic Approaches to VPC Semantics

Several studies have addressed VPCs from a cognitive linguistic standpoint (Lindner, 1981; Morgan, 1997; Hampe, 2000). Here we choose to concentrate on the work of Lindner.

Lindner gives an analysis of many VPCs which use the particles *out* and *up*, and then classifies these VPCs according to the meaning contributed by their particle. In doing so, Lindner gives an account for the semantics of many VPCs which would be considered idiomatic in terms of the analysis given by Jackendoff. Although Lindner's work looks at VPCs with both *out* and *up*, here the focus will be on VPCs with *up*.

Lindner's analysis is grounded in cognitive grammar (previously known as space grammar) which is described by Langacker (1987). Three key terms from cognitive grammar are trajector, landmark and schema.

Trajector (TR) The object which is conceptually foregrounded.

Landmark (LM) The object against which the TR is foregrounded.

Schema An abstract conceptualization of an experience. Here we focus on schemas depicting a TR, LM and their relationship in both the initial configuration and the final configuration communicated by some expression.

These concepts can be used in analyzing the semantics of a VPC since the semantic contribution of a particle corresponds to a schema. For example, in sentence (34), the TR is the balloon and the LM is the ground from which the balloon is moving away.

(34) The balloon was carried up by the wind.

The schema describing the semantic contribution of the particle in the above sentence is shown in Figure 2.1, which illustrates the relationship between the TR and LM in the initial and final configurations. Lindner identifies four schemas corresponding to senses of the particle *up*, each of which is described in turn below.

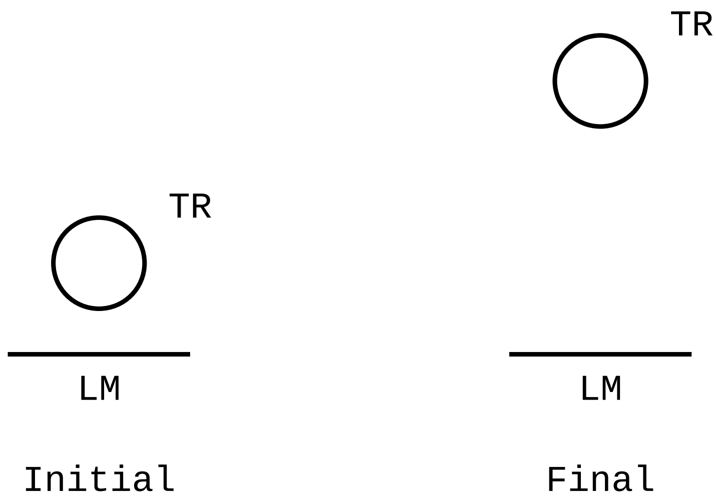

Figure 2.1: Schema for Vert-*up*.

Vertical *up* (Vert-*up*)

In this schema (shown in Figure 2.1), the TR moves away from the LM in the direction of increase along a vertically oriented axis. This includes prototypical spatial upward movement such as that in sentence (34), as well as upward movement along an abstract vertical axis as in sentence (35).

(35) The price of gas jumped up on Tuesday.

In Lindner's analysis, this sense also includes extensions of upward movement where a vertical path or posture is still salient. Note that in some of these senses, the notion of verticality is metaphorical; the contribution of such senses to a VPC might not be considered compositional in a traditional analysis. Some of the most common sense extensions are given below, with a brief justification as to why verticality is still salient.

- ***Up* as a path to increased salience in viewer's perceptual field.** Objects which are spatially high are generally easier to perceive.

  Examples: *crop up*, *dish up*, *show up*, *spring up*, *strike up*, *whip up*

- ***Up* as a path into mental field.** This is similar to the previous sense except that here *up* codes a path for mental objects as opposed to physical objects.

  Examples: *call up*, *come up (with)*, *dream up*, *dredge up*, *think up*

- ***Up* as a path into possession.** The prototypical way of acquiring an object is to raise it to hand-level.

  Examples: *grab up*, *pick up*, *snatch up*

- ***Up* as a path into a state of activity.** The prototypical way of bringing an object into an active state is to bring it into an erect posture.

  Examples: *fire up*, *gear up*, *get up*, *prick up*, *set up*, *start up*, *wake up*

- ***Up* as a path out of possession.** Objects which are spatially high are out of the range of possession.

  Examples: *fork up*, *give up*, *pass up*

- ***Up* as a path into a state of inactivity.** Objects which are spatially high are in a typical place of storage and therefore not in an active state.

  Examples: *hang up*, *lay up*

Goal-oriented *up* (Goal-*up*)

Goal-oriented *up* is characterized by the TR approaching the LM, which represents a goal, but the movement is not necessarily oriented along a vertical axis (see Figure 2.2). Prototypical examples of this sense are given below.

(36) The bus <u>drew up</u> to the stop.

(37) He <u>walked up</u> to the bar.

This sense also includes extensions into the domain of social interaction as in sentences 38 and 39, and into the domain of time as in sentences 40 and 41.

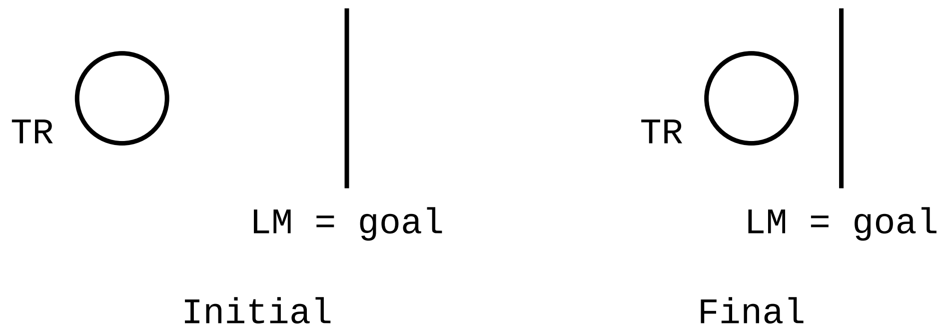

Figure 2.2: Schema for Goal-*up*.

(38) He always tries to <u>kiss up</u> to his teacher.

(39) She <u>sucks up</u> to her boss in the hope of a promotion.

(40) The deadline is <u>coming up</u> quickly.

(41) We <u>moved</u> the meeting <u>up</u> to Monday.

Completive *up* (Cmpl-*up*)

Completive *up* is a sub-sense of Goal-*up* in which the goal represents an action being done to completion. This sense shares its schema with Goal-*up* (Figure 2.2), but it is considered as a separate sense since it is very frequent and corresponds to uses of *up* as an aspectual marker. Examples of Cmpl-*up* are given below.

(42) <u>Clean up</u> your room!

(43) Suzy <u>drank up</u> all her milk.

(44) I <u>filled up</u> the car.

Reflexive *up* (Refl-*up*)

Reflexive *up* is a sub-sense of Goal-*up* in which the sub-parts of the TR are approaching each other. The schema for Refl-*up* is shown in Figure 2.3; it is unique in that the TR

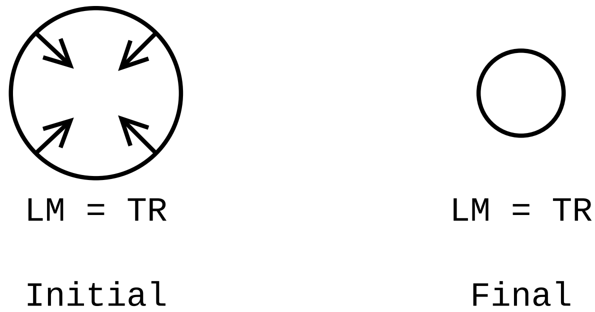

Figure 2.3: Schema for Refl-*up*.

and LM are the same object. Examples of Refl-*up* are given below.

(45) The CEO <u>bottled up</u> her anger until she burst.

(46) He <u>crumpled up</u> the piece of paper and threw it out.

(47) <u>Tie up</u> your skates!

Lindner (1982) notes that the opposite of *up* should be *down*, but gives examples, such as the following, in which this is not the case.

(48) <u>Roll out</u> the carpet and then <u>roll</u> it <u>up</u>.

(49) He <u>crumpled up</u> the letter and then <u>smoothed</u> it <u>out</u>.

Lindner uses the differing schemas for *up* (and *out*) to account for this phenomenon.

Structure of senses

Hierarchical relations among schemas can be shown in a "schematic network" (Langacker, 1987). The relationship between Vert-*up* and Goal-*up* (and the latter's sub-senses) is difficult to characterize in such a hierarchy, particularly because it is not clear whether Vert-*up* and Goal-*up* are both sub-senses of another schema. Therefore, we choose to represent the senses of *up* in a simplified schematic network, shown in Figure 2.4, in which we connect more similar senses with shorter edges.

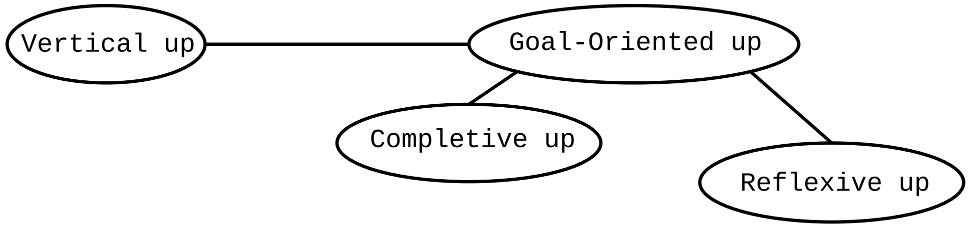

Figure 2.4: Simplified schematic network for *up*.

Limitations of Lindner's analysis

One major limitation of Lindner's analysis is that, in some cases, Vert-*up* and Goal-*up* (and the sub-senses of the latter) overlap. For example, in sentence (50), *up* could be considered to be Cmpl-*up* since the screw is being tightened completely; however, it could also be considered to be Vert-*up* since the level of tightness of the screw is being increased.

(50) John tightened up the screw.

This is an issue for any sense of *up* in which the achievement of a goal-state involves the increase of some object or property along a physical or abstract vertical axis. In her analysis, Lindner claims to have classified VPCs according to which sense of *up* is more salient.

A second limitation of Lindner's analysis is that she discusses ways in which senses of *up* may be extended to other domains, but for the most part, does not discuss the role of metaphor in these extensions. Lakoff and Johnson's (1980) conceptual metaphor theory claims that metaphor is not limited to language, but rather exists in, and provides structure for, our conceptual system. Lindner's analysis could be strengthened by accounting for these extensions of senses of *up* in terms of a system of metaphors, such as that described by Lakoff and Johnson. This type of approach is taken by Morgan (1997) in her analysis of VPCs using the particle *out*.

Chapter 3

Related Work

Studies in the computational linguistic community have examined VPCs, and the related areas of the semantic classification of verbs and preposition semantics.

3.1 Research on VPCs

Work on VPCs has focused on three issues: the identification of VPCs in text, their compositionality and their productivity.

3.1.1 Identification of VPCs

Several studies have focused on the identification of non-compositional terms (Melamed, 1997; Lin, 1999). In this section we consider work on identifying instances of VPCs, which are often non-compositional, in text.

Since the interpretation of the use of a verb and particle/preposition is ambiguous between a VPC and a verb followed by a prepositional phrase, the task of identifying VPCs in text is not trivial. Furthermore, since a transitive VPC may occur in the split construction with an arbitrary number of words between the verb and particle, n-gram based collocation identification techniques are limited (Baldwin, 2005a). Baldwin and Villavicencio (2002) examine three methods for extracting VPCs from text: a part-of-speech–based method, a chunk-based method, and a chunk grammar-based method. They evaluate their system on the Wall Street Journal corpus and find the chunk grammar-based method to perform best. They combine these three extraction techniques and augment the combined classifier with several linguistically motivated features, such as the frequency of the deverbal noun form of the VPC and the length of the verb in the VPC, to achieve an F-score of 0.865 which is better than that of any of the extraction methods used individually.

Baldwin (2005a) builds on the work of Baldwin and Villavicencio (2002), and considers the more difficult task of extracting VPCs according to their valence—i.e., extracting transitive or intransitive VPCs. He builds supervised classifiers which incorporate the three basic methods of VPC identification described by Baldwin and Villavicencio (2002), as well as a parser-based method. Again, the basic features are combined, and this combined classifier achieves F-scores of 0.969, 0.749 and 0.897 on valence under-specified, intransitive and transitive VPC extraction, respectively, from the BNC.

In other work on VPC identification, Blaheta and Johnson (2001) automatically identify verb + particle/preposition compounds. One limitation of their study is that it does not make the distinction between VPCs and prepositional verbs. Baldwin (2005b) examines the related task of automatically extracting English prepositional verb types from text.

3.1.2 VPC Compositionality

Several studies have attacked the issue of VPC compositionality. In one of the first such studies, McCarthy et al. (2003) attempt to place the compositionality of a VPC on a scale of 0 to 10—0 being completely non-compositional, 10 completely compositional. Guided by the intuition that a compositional VPC will be semantically similar to its base verb, they use an automatically created thesaurus to find the nearest neighbours of each VPC and its corresponding simplex verb (i.e., the base verb of the VPC). They examine several measures of overlap between neighbours, and find that a measure which takes into account the number of neighbours of the VPC which use the same particle as the VPC minus the number of simplex neighbours having the same particle as the VPC performs best. The intuition given for this measure is that when an expression is compositional, the particle is making a contribution to its meaning. Subtracting the number of simplex neighbours having the same particle as the VPC prevents VPCs whose simplex neighbours also include VPCs with this particle from being given larger compositionality scores. This measure gives a Spearman rank-order correlation coefficient of 0.49 with gold-standard human judgements of VPC compositionality. One limitation of this work is that it does not consider the extent to which the verb and particle contribute their semantics separately.

Bannard (2002) performs one of the first studies of VPC compositionality which distinguishes the semantic contribution of the particle from that of the verb. Some of the limitations of this initial study are addressed by Bannard et al. (2003), who consider the following four binary classification tasks involving the compositionality of VPCs.

1. Both the verb and particle contribute their simplex meaning.[1]

2. Either the verb or particle (or both) contributes its simplex meaning.

3. The verb contributes its simplex meaning.

4. The particle contributes its simplex meaning.

These tasks are interesting in that they consider the semantic contribution of the verb and particle separately; however, they are limited in that they are binary. In some cases, it can be difficult for a human judge to make a binary distinction as to whether the simplex meaning of a verb or particle is contributed in a given VPC. The gold-standard data for

[1] The simplex meaning of a particle is its basic directional sense.

this study, which consists of the compositionality judgements of twenty-six non-experts for forty VPCs, supports this. For example, in the case of *pay off*, eleven judges thought the verb's meaning was being contributed while twelve did not (the other judges gave "don't know" responses). Similarly, thirteen judges believed that *in* is implied by *throw in* while twelve did not. Bannard et al. experiment with four distributional measures of VPC compositionality which are based on the idea that non-compositional expressions are distributionally different from expressions formed by substituting semantically similar terms for the original expression's component words. However, their measures do not generally perform better than the baseline of assigning the most common class (i.e., compositional or non-compositional).

Bannard (2005) builds on the work of Bannard et al. (2003), and examines the extent to which both the verb and particle contribute their meaning in a VPC. The gold-standard data for this study is improved from that of Bannard et al. (2003); two judges with expert linguistic knowledge are also employed in addition to using non-expert judges. The expert judges were asked to rate their confidence that the verb and particle individually contribute their simplex meanings in a VPC on a scale of 1–7. The non-expert judges were asked to perform a binary classification of the simplex meaning contribution of the verb and particle. The non-expert judgements were converted to a compositionality score by dividing the total number of positive judgements by the total number of judgements. For each VPC type, Bannard creates a feature vector by counting the frequency of occurrence of all words within a small window to the left and right of the verb in each instance of that VPC across a corpus. Bannard similarly creates a feature vector for each simplex verb. He then uses the cosine of two feature vectors as a measure of their similarity. When applied to the feature vectors for a VPC and its corresponding simplex verb, this similarity measure gives a compositionality score. This score is found to correlate significantly with the compositionality judgements for the verbs, but not for the particles. Although the correlation for the verbs is significant, the strength of the correlation is very weak. The strongest correlation achieved has a Spearman rank-order correlation coefficient of just 0.25, which implies that the automatic compositionality score accounts for just 6.25% of the variation in the human judgements.

Patrick and Fletcher (2005) examine the compositionality of VPC tokens as opposed to VPC types as in the previously discussed studies. Patrick and Fletcher recognize the need for a more fine-grained classification of VPC semantics; however, in their study they only make a 3-way distinction of compositional, non-compositional and not a VPC. The third category is important since a given candidate VPC token may actually be a verb followed by a prepositional phrase. Patrick and Fletcher train a classifier using a simple set of features, which includes the particle in the VPC, the number of words occurring between the verb and particle, and the transitivity of the VPC. They achieve their best results when they exploit the fact that the verb in a VPC is usually of Germanic origin by including the last three letters of the verb as a feature. Using this set of features they achieve F-scores of 0.674, 0.633 and 0.551, respectively, on their three classes of non-compositional, compositional and not a VPC. Patrick and Fletcher also consider as a feature whether the arguments of a VPC are metaphorical. They annotate the arguments of VPCs as literal or metaphorical; however, in doing so they treat metaphoricity as a binary property, when in fact arguments may be metaphorical to varying degrees. They find that metaphorical arguments are relatively infrequent, but are able to show that if the ratio of VPCs in the training data which have metaphorical arguments to those which do not is sufficiently high, using the metaphoricity of the arguments of a VPC as a feature may improve classification accuracy. However, since they must create a corpus with an appropriate ratio of VPCs with metaphorical to non-metaphorical arguments, they are prevented from directly comparing results using features which capture the metaphoricity of arguments to results using the original set of features.

Uchiyama et al. (2005) examine the semantics of Japanese Compound Verbs (JCVs), which are a type of MWE that is composed of two verbs and bears some similarity to

English VPCs. In particular, the sense of the second verb (V2) can be either aspectual, spatial or adverbial, and the semantics of a given V2 in JCVs may be ambiguous between these senses. Uchiyama et al. seek to classify JCV tokens according to the sense contributed by their V2. They compare two methods for doing this: a statistical method and a rule-based method. In the statistical method, a feature vector is formed for each JCV j by concatenating the senses of the V2 in JCVs which use one of the component verbs of j. A classifier is trained using these features to perform word sense discrimination—i.e., reduce the number of candidate senses for the V2 in a JCV. Then from the constrained set of senses for the V2 in a JCV, the most frequent sense is returned. This method is evaluated individually on each class (i.e., aspectual, spatial and adverbial) and achieves an average accuracy of 90%. This is substantially better than the baseline of assigning the most frequent class, which has an average accuracy of 68%. The good performance is attributed to the fact that most JCVs are monosemous. In the rule-based method, a set of semantic and syntactic rules is manually developed to determine the semantic class of the V2 in a JCV. This method achieves an accuracy of 95%, but is not currently automatic—it requires a human to apply the rules.

3.1.3 VPC Productivity

Villavicencio (2005) builds on her previous work (Villavicencio, 2003) which explores the productivity of VPCs. Villavicencio notes that the number of VPCs is "constantly growing" and that the coverage of VPCs varies between existing lexical resources. Villavicencio explores a way to automatically expand the coverage of VPCs in a lexicon. Her work is based on the observation that VPCs tend to be productive across semantic classes of verbs. For example, she notes that some verbs of cooking, such as *bake*, *cook*, *fry* and *broil*, can all combine with the particle *up* to form a VPC. However some cooking verbs, such as *saute*, are less acceptable when combined with *up*. In this study, Villavicencio generates candidate VPCs by combining verbs from Levin's (1993) classes with the particles *away*, *down*, *in*, *on*, *out* and *up*. Candidate combinations which are not found in a set of lexical resources are searched for in the World Wide Web using the Google search engine. Villavicencio uses a restricted search pattern which requires that one of a small set of prepositions occurs after the particle in a VPC to prevent matching uses of a verb followed by a prepositional phrase. For example, when looking for evidence for *walk up*, Villavicencio searches for *walk up from*, to avoid the possibility that *up* is the head of a prepositional phrase. VPCs which are found in the search are considered to be valid VPCs. Villavicencio finds that verbs of similar semantic classes tend to form VPCs with similar sets of particles. One drawback of this study is that the evaluation is somewhat unsatisfactory in that it does not actually verify, using any sort of human judgments, that the automatically generated VPCs are valid.

3.2 Verb Classification

Work on verb classification—automatically assigning a verb to a semantic class—is crucial to the task of particle sense classification, since the meaning of a VPC appears to be related to the meaning of its base verb. Here we examine some of the research in this area.

Merlo and Stevenson (2001) classify English verbs into three classes according to the thematic role assigned to their arguments. For classification they use five simple features which can easily be extracted from text pre-processed using standard tools such as a part-of-speech tagger and parser. They report an accuracy of 69.8% on a task which has a baseline of 34% and an upper bound of 86.5%. One drawback to this work is that the features they use are manually designed specifically for this application, and might not be useful for other tasks. Joanis and Stevenson (2003) and Joanis et al. (2006) address this shortcoming by devising a general feature space, based on verb alternations identified by Levin (1993), to classify verbs according to their semantic class. They evaluate their features on several classification tasks and report reductions in error rate of 49–90%. They find that their features which capture the syntactic slots in which the arguments and adjuncts of a verb are expressed, which are heavily relied on in this thesis, are particularly useful for verb classification.

Schulte im Walde (2003) addresses the task of clustering German verbs according to their semantic class. To do this she uses a set of features based on sub-categorisation frames and selectional preferences. Schulte im Walde's results are better than the baseline of randomly assigning verbs to classes, but are well below the upper bound performance achieved by manually assigning verbs to clusters. Schulte im Walde (2005) applies the same features used by Schulte im Walde (2003) to the task of identifying semantic nearest neighbours of German particle verbs, a somewhat related phenomenon to English VPCs. Schulte im Walde claims that determining the semantic nearest neighbour of a particle verb is a first-step towards being able to determine its compositionality. In particular, more-compositional particle verbs will be more similar to their simplex verbs (Schulte im Walde, 2004).

3.3 Preposition Semantics

There is a clear relationship between preposition and particle semantics. Since there has been little work done in the computational linguistic community addressing the issue of determining the semantic contribution of a particle, we also examine the work done on determining the semantic contribution of a preposition.

Some work has addressed the issue of preposition semantics in general. O'Hara and Wiebe (2003) seek to classify a use of a preposition according to the semantic role conveyed. They use standard word-sense disambiguation features such as the words and parts-of-speech surrounding the target word to be disambiguated. In addition to words which occur near the target, they also consider WordNet (Fellbaum, 1998) hypernyms of these words, to capture the semantic categories which occur near the target. O'Hara and Wiebe report an accuracy of 86.1% for identifying the more coarse-grained Penn Tree-Bank (Marcus et al., 1994) semantic roles and 49.4% for the more fine-grained FrameNet (Fillmore et al., 2001) semantic roles. These results are a significant improvement over the baseline of assigning the most frequent class, which is 48.0% and 14.9% for the Penn TreeBank and FrameNet roles respectively. In a very different study of the semantics of prepositions, Litkowski (2002) uses a dictionary to create a directed graph where vertices represent prepositions and edges represent hypernymic relations between preposition senses. Litkowski describes how such a graph can be used to find a basic set of prepositions from which other prepositions may be derived.

Baldwin (2006) investigates the use of distributional measures of similarity for prepositions and particles; such measures have previously been applied to parts-of-speech such as nouns and verbs, but not prepositions and particles. Baldwin uses a similarity score based on Latent Semantic Analysis (LSA, Deerwester et al., 1990) and compares it to similarity measures based on lexical conceptual structures (LCS, Dorr, 2001) and a thesaurus. He finds a modest correlation between his LSA-based score and the LCS-based measure. Baldwin finds a good correlation between his score and the thesaurus-based measure, but only for particles and only when information about the semantic class of the VPC in which the particle participates is included.

Some work has also examined the more fine-grained semantics of a particular preposition. Alam (2004) identifies senses of *over* and creates two sets of features: the first based on complements of prepositional phrases with *over*, the second on heads governing such prepositional phrases. From these sets of features, Alam builds two decision trees, one using the complement features, the other using the head features, to disambiguate uses of *over*. Alam manually applies the decision trees to 295 uses of *over* as a preposition, and reports an accuracy of 93.6%, but does not give a baseline for this task. Alam notes that automatic application of the decision trees would require recognition of prepositional uses of *over*, and identification of the semantic class of the head and complement. Boonthum et al. (2005) analyze the senses of *with* from the LCS preposition database (Dorr, 2001), and come up with a set of rules based on the complements and heads which are used with this preposition in its various senses. These rules are applied to instances of *with* to identify a set of its possible senses. Boonthum et al. evaluate their algorithm on eight sentences for each of fifteen manually selected verbs which are used as heads governing *with*. Out of the 120 test sentences, their system returns exactly the correct sense of *with* twenty-six times, and a set of senses of *with* containing the correct sense sixty times.

There exist lexical resources for verbs and nouns such as VerbNet and WordNet, but until recently, there have been no similar resources for prepositions. Two projects which are working to fill this gap are The Preposition Project (Litkowski, 2005) and PrepNet (Saint-Dizier, 2005). These resources are not suitable as the basis for the sense classes in this thesis because they do not address the range of metaphorical extensions that a preposition or particle can take on; however, future work may enable larger scale studies of the type needed to adequately address VPC semantics.

Chapter 4

Computational Models of Particle Semantics

In this study, we aim to classify VPCs using the particle *up* according to the semantics of their particle. However, we would like to develop a *general* set of features that captures distinctions among the different senses of any particle, not just *up*, and which may be used for sense classification of other particles in the future. In this chapter we first describe the features used for sense classification of particles, and then examine the classes, which correspond to senses of *up*, that serve as the basis for our classification task.

4.1 Features Used in Classification

We develop two sets of features, linguistic features and word co-occurrence features, which differ in terms of their motivating principles. Each set of features is described in turn in the following subsections.

4.1.1 Linguistic Features

The linguistic features are composed of the slot, adverb, nominal, and particle features. These features are motivated by specific semantic and syntactic properties of verbs and VPCs.

Slot Features

We observe that VPCs which are formed from verbs of the same semantic class, and which use a common particle, often draw on the same meaning of that particle. As evidence of this, consider the VPCs *drink up*, *eat up* and *gobble up*; all of these draw on the completion sense of *up*. As another example, each of the VPCs *puff out*, *spread out* and *stretch out* draws on the extension sense of *out*. Villavicencio (2005) has noted that verbs of the same semantic class will tend to form VPCs with similar sets of particles. Here we further hypothesize, from observations such as those noted above, that the semantic contribution of a particle when combined with a given verb is related to the semantics of that verb. That is, the particle contributes the same meaning when combining with any of a semantic class of verbs. The prevalence of the aforementioned patterns suggests that features which have been shown to be effective for the semantic classification of verbs may be useful for our task of semantic classification of particles.

We adopt simple syntactic "slot" features which have been successfully used in the automatic semantic classification of verbs (Joanis and Stevenson, 2003; Joanis et al., 2006). These features are motivated by the fact that semantic properties of a verb are reflected in the syntactic expression of the participants in the event the verb describes (Levin, 1993). The syntactic slots are subject, direct and indirect object, and object of a preposition, the latter distinguished by the identity of the preposition (i.e., we consider arguments of different prepositions separately). The slot features encode the relative frequencies of the syntactic slots that the arguments and adjuncts of a verb appear in. We calculate the slot features separately over the following three contexts for each target expression (a VPC using *up* in our experiments):

All uses of the base verb of the target expression. To capture the semantics of the base verb of the target VPC, and thus the semantics of the target.

All uses of the target expression. To directly learn about the semantics of the target VPC.

All uses of the base verb of the target expression in a VPC with any of a set of high-frequency particles. To gain information about the semantics of the base verb of the target VPC when used in VPCs in general.

These three sets of features together form the slot features for a target VPC to be classified.

Adverb Features

Another indication of the semantic class of a verb is its pattern of co-occurrence with adverbs; verbs of similar semantic classes will tend to occur with similar sets of adverbs. We therefore hypothesize that verbs with similar patterns of co-occurrence with adverbs will behave similarly semantically when used in a VPC. For each target expression, we count the relative frequency of occurrence of each of a set of high-frequency adverbs in each of the three contexts described above for the slot features. Our adverb features are similar to features used by Joanis et al. (2006).

Nominal Features

Lindner (1981) notes that denominal verbs which describe applying or providing one object to another tend to combine with *up* in similar ways. Consider the following examples, which are taken from Lindner (1981):

(51) The bay silted up.

(52) Don't doodle up your handout.

(53) John saddled up his horse and rode out of town.

(54) GM tooled up their new factory.

In sentences 51 and 52, silt and doodles are being put into the bay and onto the handout respectively, while in sentences 53 and 54, a horse and a factory are being provided with a saddle and tools. Furthermore, in the second two examples, the horse and factory are being brought into a state of readiness for some action as a result of being provided with these objects. To capture these trends as a feature, we count the relative frequency of occurrence of any noun which has the same form as the base verb of the target expression, in a manner similar to Joanis et al. (2006).

Particle Features

Two types of features are motivated by properties specific to the semantics and syntax of particles and VPCs.

First, Wurmbrand (2000) notes that compositional particle verbs in German (a somewhat related phenomenon to English VPCs) allow the replacement of their particle with semantically similar particles. This property may also be true of English VPCs. For example, both *bring up* and *move up* are compositional and their base verbs form VPCs with the particles *back*, *down*, *in* and *out* as well. In contrast, *muck up* is relatively non-compositional, and does not form a VPC when combined with any of the particles *back*, *down*, *in* or *out*.

Based on Wurmbrand's observations about German particle verbs and similar patterns in English VPCs, we hypothesize further that when a verb combines with a particle such as *up* in a particular sense, the pattern of usage of that verb in VPCs using all other particles may be indicative of the sense of the target particle (in this case *up*) when combined with that verb. To reflect this hypothesis, we count the relative frequency of any occurrence of the base verb of the target expression used in a VPC with each of a set of high-frequency particles, described in detail in Section 5.2.2.

Second, as noted in Section 2.1, one of the striking syntactic properties of VPCs is that they can often occur in either the joined construction or the split construction. Bolinger (1971) notes that VPCs which are idiomatic according to his analysis may show a preference for occurring in the joined construction. Bolinger argues that the final position receives "semantic focus" and therefore a particle that has little meaning is unlikely to occur in this position. One situation in which this is particularly evident is when the object of a transitive VPC is a clause beginning with a *wh*-word. Sentences 55 and 56, taken from Bolinger (1971), demonstrate that a so-called idiomatic VPC is not valid in the split construction, while sentences 57 and 58 show that a literal VPC is acceptable in both the split and joined constructions.

(55) I can't <u>make out</u> who it is.

(56) *I can't <u>make</u> who it is <u>out</u>.

(57) Regretfully, he <u>gave back</u> what he had found.[1]

(58) Regretfully, he <u>gave</u> what he had found <u>back</u>.

Bolinger also gives the following examples which show that when a literal VPC undergoes nominalization it may appear in both the joined and split constructions (sentences 59 and 60), while idiomatic VPCs are generally only acceptable in the joined construction (sentences 61 and 62).

(59) His <u>throwing up</u> of the ball was stupid.

(60) His <u>throwing</u> of the ball <u>up</u> was stupid.

[1] This example is not given in Bolinger (1971); however, it is included here to contrast with sentence (58).

(61) His throwing up of his dinner was stupid.

(62) *His throwing of his dinner up was stupid.

Recall from Section 2.1 that many VPCs allow the insertion of an adverb between the verb and particle. Bolinger notes that idiomatic VPCs resist insertion of adverbs more than literal VPCs. In the following examples, taken from Bolinger, the first contains a literal VPC which allows insertion of an adverb, while the second contains an idiomatic VPC which does not.

(63) They clattered noisily on.

(64) *He caught quickly on.

In this study we assume that all VPCs are compositional; however, we also believe that they lie on a continuum from literal to idiomatic. In the spirit of Bolinger's analysis, we hypothesize that more idiomatic VPCs will tend to favour the joined construction, while more literal VPCs will be more flexible. To encode this as a feature, we calculate the relative frequency of the verb co-occurring with the particle *up* with each of 0–5 words between the verb and *up*, reflecting varying degrees of verb-particle separation.

4.1.2 Word Co-occurrence Features

These features differ from the linguistic features in that they are not motivated by semantic and syntactic properties specific to verbs or VPCs. Instead these are general context features, in the form of word co-occurrence frequency vectors, which have been used in numerous approaches to determining the semantics of a target word. However, it is important to note that unlike the task of word sense disambiguation, which examines the context of a target word *token* to be disambiguated, here we are looking at aggregate contexts across all instances of a target VPC, in order to perform *type* classification (i.e., classification of the semantics of the particle used in this VPC expression overall).

We adopt very simple word co-occurrence features (WCFs), calculated as the frequency of any (non-stoplist) word within a small window to the left and right of the target expression. We noted above that the target particle semantics is related both to the semantics of the verb it co-occurs with, and to the occurrence of the verb across VPCs with different particles. Thus we not only calculate the WCFs of the target VPC (a given verb used with the particle *up*), but also the WCFs of the verb itself, and the verb used in a VPC with any of the high-frequency particles. These WCFs give us a very general means for determining semantics, whose performance we can contrast with our linguistic features.

4.2 The Sense Classes Used for Our Study

We classify target VPCs according to which of the senses of *up*, described in Section 2.2.2 and repeated below for convenience, is contributed to the expression.

- Vert-*up*
- Goal-*up*
- Cmpl-*up*
- Refl-*up*

For example, the expressions *jump up* and *pick up* are designated as being in the class Vert-*up* since *up* in each of these VPCs has the vertical sense, while *clean up* and *drink up* are designated as being in the class Cmpl-*up* since *up* in these expressions has the completive sense.

Recall that the senses of *up* can be organized into a schematic network, as discussed in Section 2.2.2. One of the main motivations for basing our sense classes on a cognitive linguistic analysis is the structure of the schematic network of senses; combining closely related senses allows us to alter the granularity of our classification in a linguistically motivated fashion. For example, Cmpl-*up* or Refl-*up* could be merged with Goal-*up*, since they are both sub-senses of, and semantically similar to, Goal-*up*, as shown in Figure 2.4. Thus we can explore the effect of different sense granularities on classification.

Chapter 5

Materials and Methods

5.1 Experimental Expressions

We created a list of English VPCs using the particle *up* based on a publicly available list of VPCs (McIntyre, 2001) and a list of VPCs created by two human judges, both of whom were native English speakers. The judges then independently rated each VPC as acceptable or not, and any VPC which either judge thought to be unacceptable was discarded. The final list contained 389 VPCs.

The VPCs in this list are split into three frequency ranges according to how often their base verb occurs with any verb part-of-speech tag in the British National Corpus (BNC, Burnard, 2000), a corpus of approximately 100M words. Base verb frequency, as opposed to VPC frequency, is used for splitting the expressions into frequency ranges, since many of the features used in classification, described in Section 4.1, depend on the use of a verb. Furthermore, the frequency of a VPC is only approximate, since automatic VPC identification is challenging. Training, verification, and test sets of sixty VPCs each are formed by randomly selecting VPCs from the frequency ranges, such that the proportion of VPCs in each frequency range is the same in each dataset. The frequency ranges, number of VPCs in each frequency range, and number of VPCs per dataset in each

Frequency Range of Base Verb	Total #VPCs	#VPCs per Dataset
freq < 100	42	0
100 <= *freq* < 1000	143	25
1000 <= *freq* < 5000	101	17
5000 <= *freq*	103	18

Table 5.1: Number of VPCs in frequency range of base verb.

frequency range are given in Table 5.1.

Each VPC in each dataset is independently annotated by each judge according to which of the four senses of *up* identified in Section 4.2 is contributed by its particle. The observed inter-annotator agreement for this task is 0.80 for each dataset. The unweighted kappa scores are 0.73, 0.64 and 0.55, for the training, verification and test sets respectively. After this initial round of annotation, the judges discussed VPCs on which they disagreed, and together determined a consensus classification.

It is important to note that VPCs may be ambiguous with respect to their particle sense; for example, *come up* may be used with *up* in the vertical sense as in sentence (65), or in the goal-oriented sense as in sentence (66).

(65) The sun <u>came up</u>.

(66) The deadline is <u>coming up</u>.

However, following the type-based approaches to VPCs of McCarthy et al. (2003) and Bannard et al. (2003), we simplify our task by having the judges assign each VPC to a single sense class for *up* according to their assessment of its predominant usage. Furthermore, the use of *up* in a VPC may draw on multiple senses of this particle, as discussed in Section 2.2.2. In such cases, the judges were asked to choose the sense which they thought was most salient. The particle sense contribution judgements are given in full in Appendix A.

5.2 Calculation of the Features

In the following subsections we describe how our features are calculated for each VPC using counts extracted from the BNC. Since we calculate the slot, adverb and word co-occurrence features over three different contexts, as described in Section 4.1.1, the term *target expression* will refer to the following: a VPC using *up*, the base verb of the VPC, and the base verb of the VPC used in a VPC with any of the high-frequency particles, defined below.

5.2.1 VPC Identification

The calculation of many of our features requires the identification of VPCs in text. We identify VPCs using a simple heuristic based on part-of-speech (POS) tags, similar to the POS-based method used by Baldwin (2005a). According to our heuristic, a use of a verb v is considered part of a VPC if it occurs with a particle p (any word tagged AVP) within a six-word window to the right with neither another verb nor another particle occurring between v and p. We feel that six words is a reasonable window size since, as discussed in Section 2.1, VPCs with heavy noun-phrase complements tend to occur in the joined construction, and a larger window size would give noisier results. Over a random sample of 113 VPCs extracted using our heuristic, we find the precision to be 88%, somewhat below the performance of Baldwin's (2005a) best extraction method, indicating potential room for improvement. (The recall cannot be estimated since we do not know the true number of VPCs in the BNC.)

5.2.2 Linguistic Feature Calculation

The features motivated by syntactic and semantic properties of verbs and VPCs (the slot, adverb, nominal and particle features) are calculated using a modified version of the ExtractVerb software provided by Joanis et al. (2006), which runs over the BNC. The slot and adverb features require that the corpus be pre-processed using the Cass chunker (Abney, 1991).

Slot Features

The slot features capture information about the syntactic slots with which a target expression occurs. We count the relative frequency of the target expression with each of the following slots (see Joanis et al. (2006) for details):

- subject
- subject of a transitive verb
- subject of an intransitive verb
- object
- direct object
- indirect object
- prepositional phrase (identified by the particular preposition)

We also count the relative frequency of occurrence of the target expression with prepositional phrases headed by each of a set of high-frequency prepositions, listed in Table 5.2. Joanis et al. (2006) define the high-frequency prepositions to be those which occur more than 10 000 times in the BNC. We also include the preposition *up*, even though it does not meet the frequency cut-off, since it is homonymous with the particle which we are investigating, and there is a clear relationship between preposition and particle semantics. Joanis et al. permit spelling variation for some prepositions; these prepositions are listed in Table 5.3 with their alternative spelling.

To capture information about less frequent prepositions, Joanis et al. (2006) group prepositions that are similar in meaning or expected to be used similarly. We follow their

about	as	by	like	rather than	until
above	as well as	despite	near	round	up
according to	at	during	of	since	upon
across	away from	for	off	such as	up to
after	because of	from	on	through	with
against	before	in	outside	throughout	within
along	behind	including	out of	to	without
among	between	into	over	towards	
around	beyond	in terms of	per	under	

Table 5.2: High-frequency prepositions, taken from Joanis (2002), with the exception of *up*.

Preposition	Alternative Spelling
about	'bout
for	fer
of	o'
over	o'er
with	wi'

Table 5.3: Prepositions for which an alternative spelling is allowed, taken from Joanis (2002).

approach, and count the relative frequency of occurrence of the target expression with any preposition from each of a set of nineteen groups of prepositions, which are given in Table 5.4.

The slot features also capture a very limited amount of information about the pattern of occurrence of the target expression with adverbs. In particular, we follow Joanis et al. (2006) and count the relative frequency of occurrence of an adverb following the target expression. Levin (1993) notes that in some cases a transitive verb may be used intransitively with a modifier. Therefore we again follow Joanis et al., and capture this pattern by counting the relative frequency of an adverb following the target expression when it is used intransitively.

Adverb Features

The adverb features expand on the limited information which the slot features provide about the target expression's pattern of occurrence with adverbs. Here we count the relative frequency of occurrence of each of the adverbs in Table 5.5 in a verb chunk with the target expression, and each of the adverbs in Table 5.6 occurring after a verb chunk containing the target expression.

Nominal Feature

The nominal feature is calculated by counting the number of times the base verb of the VPC occurs as a noun, divided by the number of times it occurs as either a noun or verb.

Particle Features

To calculate the particle features, we count the number of times the base verb of each target VPC occurs in a VPC (according to our identification heuristic) with each of a set of fifteen high-frequency particles, and then divide by the total number of times the base verb of the target VPC occurs. The high-frequency particles, given in Table 5.7,

Group Name	Prepositions in Group
above	above, on top of
add	as well as, besides, including, in addition to, in conjunction with, plus
behind	behind, in front of
between	among, amongst, amid, amidst, between, in between
cause	because of, for fear of, on account of
despite	despite, in spite of, notwithstanding
dest	into, onto, on to
during	during, throughout
except	apart from, aside from, bar, barring, but, but for, except, excepting, except for, excluding, other than, save, save for
inside	inside, outside, outside of
instead	instead of, rather than
like	like, unlike
near	adjacent to, beside, close to, near, nearer, nearest, near to, nearer to, nearest to, next to, opposite
path	across, along, around, beyond, down, past, round, through, toward, towards, up
regard	as for, as regards, as to, concerning, in regard to, in view of, pertaining to, re, regarding, with regard to, with respect to
reltime	after, before, prior to, since, till, until
source	away from, off, off of, out, out of
spatial	adjacent to, above, behind, below, beneath, beside, close to, in front of, near, nearer, nearest, near to, nearer to, nearest to, next to, on top of, opposite, outside, outside of, over, o'er, under, underneath
under	below, beneath, under, underneath

Table 5.4: Preposition groups, taken from Joanis (2002).

n't	also	just	now	still	never
always	so	already	really	even	actually
often	in order	ever	enough	probably	simply
usually	sometimes	certainly	much	thus	therefore

Table 5.5: High-frequency adverbs in verb chunk.

away	again	together	there	forward	here
now	too	so	home	just	about
more	both	aside	over	yesterday	today
not	all	more than	at all	further	at least
right	up to	either	seriously	straight	slowly
even					

Table 5.6: High-frequency adverbs following verb chunk.

are those which occur more than 100 times in the BNC with the POS tag AVP. We also count the number of times the target VPC occurs with each of 0–5 words between the verb and particle, and divide by the total number of times the target VPC occurs.

5.2.3 Word Co-occurrence Feature Calculation

To compute the word co-occurrence features (WCFs), we first determine the relative frequency of all words which occur within a five-word window to the left and right of any of the target expressions in the training data. From this list we eliminate the most frequent 1% of words as a stoplist, and then use the next n most frequent words as "feature words". For each target expression, we then calculate the relative frequency of occurrence of each feature word within the same five-word window to the left and right. We use $n = 200$ and $n = 500$ to create feature sets WCF_{200} and WCF_{500} respectively.

Particle	Frequency
up	158064
out	145706
back	75233
down	72709
on	54956
off	37751
in	34411
over	32526
about	12587
round	10895
around	10384
through	5796
along	4925
by	371
under	313

Table 5.7: High-frequency particles and their frequency in the BNC.

Sense Class	#VPCs in Sense Class		
	Train	Verification	Test
Vert-*up*	24	33	27
Goal-*up*	1	1	3
Cmpl-*up*	20	23	22
Refl-*up*	15	3	8

Table 5.8: Frequency of items in each sense class.

5.2.4 Feature Contexts

Recall from Section 4.1 that the slot, adverb and WCF features are calculated independently for three contexts, in which the target expression differs: all uses of the VPC, the base verb of the VPC, and the base verb of the VPC used in a VPC with any of the high-frequency particles.

5.3 Experimental Classes

Table 5.8 shows the distribution of senses in each dataset. Each of the training and verification sets has only one VPC corresponding to Goal-*up*. Recall from Section 2.2.2 that the network in which the senses of *up* are arranged gives linguistic motivation for combining senses. Goal-*up* shares a schema with Cmpl-*up*, and is therefore very close to it in meaning, as indicated spatially in the sense network in Figure 2.4 (page 17). We therefore merge Goal-*up* and Cmpl-*up* into a single sense, to provide more-balanced classes.

One of the goals of this study is to explore the effect of differing granularities of senses on classification. We run each experiment as both a 3-way and 2-way classification task, merging senses as shown in Tables 5.9 and 5.10 respectively. In the 3-way task, the sense classes correspond to the meanings Vert-*up*, Goal-*up* merged with Cmpl-*up* (as noted

Sense Class	#VPCs in Sense Class		
	Train	Verification	Test
Vert-*up*	24	33	27
Goal-*up* + Cmpl-*up*	21	24	25
Refl-*up*	15	3	8

Table 5.9: Frequency of items in each class for the 3-way task.

Sense Class	#VPCs in Sense Class		
	Train	Verification	Test
Vert-*up*	24	33	27
Goal-*up* + Cmpl-*up* + Refl-*up*	36	27	33

Table 5.10: Frequency of items in each class for the 2-way task.

above), and Refl-*up*. In the 2-way task, we further merge the classes corresponding to Goal-/Cmpl-*up* with that of Refl-*up*, because as illustrated in Figure 2.4 (page 17), Refl-*up* is also a sub-sense of Goal-*up*. Moreover, all three of these senses contrast with Vert-*up*, in which increase along a vertical axis is the salient property. It is worth emphasizing that the 2-way task is not simply a classification between literal and non-literal *up*—Vert-*up* includes extensions of *up* in which the increase along a vertical axis is metaphorical.

5.4 Evaluation Metrics

The variation in frequency of the sense classes of *up* across the datasets makes the true distribution of the classes difficult to estimate. Furthermore, there is no obvious informed baseline for this task. Therefore, we make the assumption that the true distribution of the classes is uniform, and use the chance accuracy $1/C$ as the baseline (where C is the number of classes—in our experiments, either 2 or 3). Given this assumption and corresponding baseline, our measure of classification accuracy should weight each class evenly. Therefore, we report the average per class accuracy, which gives equal weight to each class.

5.5 Classifier Software

Before describing the classifier employed, we describe the pre-processing which we apply to the data.

Following Joanis et al. (2006), we begin by replacing any missing value—a feature value that could not be calculated due to division by 0—by the 60% trimmed mean for that feature in the dataset in which the missing value occurs. The 60% trimmed mean is computed by eliminating the 30% highest and lowest values, and then taking the mean of the remaining values. We then eliminate any feature which takes on the same value across all data points in the training data. For each remaining feature, we calculate its 60% trimmed mean and mean-absolute-deviation in the training data. Then for each feature in both the training and testing data, we subtract the corresponding trimmed mean and divide by the appropriate mean-absolute-deviation. Finally, we take the arctan of each dataset to reduce the effects of outliers (Sarle, 2002).

For classification we use LIBSVM (Chang and Lin, 2001), an implementation of a support-vector machine (SVM). LIBSVM provides many different options for classification; following Joanis et al. (2006), we use the default parameters suggested by Hsu et al. (2003). They recommend setting the SVM type to C-SVC and the kernel type to a radial basis function. These choices require us to set two additional parameters, cost and gamma. Using the same method as Joanis et al. (2006), we perform a grid search through possible values of these parameters to find the optimal combination. For cost we consider the range $2^{17}, 2^{15}, 2^{13} ... 2^{-5}$ and for gamma $2^{3}, 2^{1}, 2^{-1} ... 2^{-17}$. For each combination of values we perform 10-fold cross-validation on the training data using ten random restarts.

Since we use a uniform baseline, as described in Section 5.4, we again follow the approach of Joanis et al. (2006) and assign a weight of $\frac{|Largest\ \ Class|}{|Class\ \ c|}$ to each class c. This has the effect of increasing the penalty for misclassification of datapoints in low-frequency classes, so that the classifier will not be biased towards higher frequency classes.

Note that our choice of accuracy measure and weighting of classes in the classifier is necessary given our assumption of a uniform random baseline. Since the accuracy values we report incorporate this weighting, these results cannot be compared to a baseline of always choosing the most frequent class.

Chapter 6

Experiments and Results

We present experimental results for both verification and unseen test data, on each set of features described in Section 4.1, used individually and in combination.

Recall from Section 4.1 that the slot features, adverb features and WCFs are calculated separately over three different contexts. Preliminary experiments on verification data indicated that combinations of contexts which include the target expression to be classified—a VPC with *up* in our experiments—gave the best results. Therefore, we perform experiments on test data calculating the slot features, adverb features and WCFs for the following combinations of contexts:

VPCs with *up* and verbs All uses of the target expression + all uses of the base verb of the target expression.

VPCs with *up* and all VPCs All uses of the target expression + all uses of the base verb of the target expression used in a VPC with any of the high-frequency particles, described in Section 5.2.2.

All contexts All uses of the target expression + all uses of the base verb of the target expression + all uses of the base verb of the target expression used in a VPC with any

Features	3-way Task		2-way Task	
	Ver	Test	Ver	Test
Particles	37	33	65	47
Slots	41	51	53	67
Slots + Particles	54	54	59	63
All Linguistic	54	50	68	63

Table 6.1: Accuracy (%) using linguistic features.

of the high-frequency particles.

We find that features which are calculated for the combination of all contexts generally perform better than features calculated for other combinations of context. Therefore, in this chapter we focus on experiments using the full combination of contexts, but also present results for experiments using other combinations of context in Appendix B. All experiments are run on both the 3-way and 2-way sense classification task, described in Section 5.3, which have a chance baseline of 33% and 50%, respectively. Tables 6.1–6.4 give results for experiments using the full combination of contexts.

6.1 Experiments Using Linguistic Features

The results for experiments using the features which are motivated by semantic and syntactic properties of verbs and VPCs are summarized in Table 6.1, and discussed in turn below.

6.1.1 Particle Features

We examine the performance of the particle features on their own, since experiments using just these features indicate the extent to which patterns of combination of a verb with particles can indicate the semantics of a specific particle when combining with that verb. The results are disappointing, with only the verification data on the 2-way task showing substantially higher accuracy than the baseline. An analysis of errors reveals no consistent explanation, suggesting that the variation may be due to small sample sizes.

6.1.2 Slot Features

Experiments using the slot features alone test whether features that tap into semantic information about a verb are sufficient to determine the appropriate sense class of a particle when that verb combines with it in a VPC. Although accuracy on the test data is well above the baseline in both the 2-way and 3-way tasks, for verification data the increase over the baseline is much less. The class corresponding to the sense Refl-*up* in the 3-way task is relatively small in both the verification and test sets, as shown in Table 5.8 (page 45). This means that a small variation in classification of these VPCs may lead to a large variation in accuracy, since our measure reports the average per class accuracy as discussed in Section 5.4. However, this is not the cause of the variation in accuracy across the verification and test sets, since the accuracy on VPCs in the sense class for Refl-*up* is similar in both datasets. Although these features show promise for our task, the variation in accuracy between verification and test data indicates the limitations of our small sample sizes.

6.1.3 Slot + Particle Features

We hypothesize that the combination of the slot features with the particle features will give an increase in performance over either set of features used individually, given that they tap into differing properties of verbs and VPCs. Although we do not find the combination of the slot and particle features to give an increase in performance in all cases, we do find that the use of these features together gives more consistent performance across verification and test data than either feature set used on its own. This indicates that the classifier is learning a more general model that is less sensitive to variation amongst the datasets. We analyze the errors made using the slot and particle features separately, and find that they tend to classify different sets of verbs incorrectly. Therefore, we conclude that these feature sets are at least somewhat complementary. By combining these complementary feature sets, the classifier is better able to generalise across different datasets.

Further examining the results, we note that in the case of verification data for the 3-way task, the accuracy increases substantially from the experiment using just the slot features. This occurs because this experiment classifies incorrectly one fewer VPC whose true sense class corresponds to Refl-*up*. These findings demonstrate that results for the 3-way task are highly sensitive to small differences in classification of VPCs in the Refl-*up* sense class, especially on verification data.

6.1.4 All Linguistic Features

We would like to see how well the classifier performs using all and only those features which are motivated by syntactic and semantic properties of verbs and VPCs. We therefore consider the full set of linguistic features—i.e., the combination of the slot, adverb, nominal and particle features. Only one experiment, the 2-way task on verification data, shows an improvement over the corresponding experiment using just the slot and particle features. A detailed analysis of the errors made by the latter classifier shows that it classifies too many verbs as the sense class corresponding to the amalgamation of the senses Goal-*up*, Cmpl-*up* and Refl-*up*. This does not occur in the experiment using all the linguistic features; however, we are unable to account for why this is the case. We conclude that the additional linguistic features capture little, if any, information which is not provided by the combination of the slot and particle features.

	3-way Task		2-way Task	
Features	Ver	Test	Ver	Test
WCF_{200}	45	42	59	51
WCF_{500}	38	34	55	48

Table 6.2: Accuracy (%) using WCFs.

6.2 Experiments Using WCFs

Our goal is to compare the more knowledge-rich linguistic features to an alternative feature set, the WCFs, which does not rely on linguistic analysis of the semantics and syntax of verbs and VPCs. Recall that we experiment with both 200 feature words, WCF_{200}, and 500 feature words, WCF_{500}, as shown in Table 6.2. For each combination of task and dataset, the accuracy using all the linguistic features is higher than that for both WCF_{200} and WCF_{500}. From these results, it appears that features based on semantic and syntactic properties of verbs and VPCs are better suited to our task than linguistically uninformed WCFs.

6.3 Experiments Combining Linguistic Features and WCFs

Although the WCFs perform worse than the linguistic features, an analysis of errors shows the two sets of features to be at least somewhat complementary, since they tend to classify different verbs incorrectly. We hypothesize that as with the slot and particle features, the different types of information provided by the linguistic features and WCFs may improve performance in combination. We perform two experiments which combine linguistic features and WCFs: the slot and particle features combined with the WCFs (Table 6.3), and all linguistic features combined with the WCFs (Table 6.4). However, contrary to our

Features	3-way Task		2-way Task	
	Ver	Test	Ver	Test
Combined_{200}	53	45	63	53
Combined_{500}	54	46	65	49

Table 6.3: Accuracy (%) combining slot and particle features with WCFs.

Features	3-way Task		2-way Task	
	Ver	Test	Ver	Test
Combined_{200}	57	45	62	55
Combined_{500}	65	47	60	50

Table 6.4: Accuracy (%) combining all linguistic features with WCFs.

hypothesis, the experiments using the combination of the two feature sets do not show these features to consistently perform better than the linguistic features on their own. This variation indicates that larger sample sizes are needed to draw firmer conclusions about the effects of using the WCFs in combination with the linguistic features.

6.4 Discussion of Results

The best performance on unseen test data for the 3-way task is achieved using the slot and particle features calculated over the combination of all contexts. The best results on test data for the 2-way task are achieved using just the slot features, also calculated for the combination of all contexts. The linguistically uninformed WCFs perform worse on their own, and do not consistently help (and in some cases hurt) the performance of the linguistic features when combined with them. We conclude then that features based on semantic and syntactic properties of verbs and VPCs are motivated for this task. Note that the features are still quite simple, and straightforward to extract from a corpus—i.e., linguistically informed does not mean expensive (although the slot features do require access to chunked text).

Interestingly, in determining the semantic nearest neighbor of German particle verbs, Schulte im Walde (2005) found that WCFs that are *restricted* to the arguments of the verb outperform simple window-based co-occurrence features. Although her task is quite different from ours, similarly restricting our WCFs may enable them to encode more linguistically-relevant information.

In our experiments using all the linguistic features, we find that for each task and dataset, calculating the slot and adverb features over all three contexts achieves the highest accuracy—the results using all linguistic features in Table 6.1 are higher than those in Table B.1 in Appendix B. This is as we expect, since the classifier is provided with the most information. However, we do not observe this trend in the experiments which use just the WCFs, the results for which are shown in Tables 6.2 and B.2. The linguistic features and WCFs are very different types of features, and it appears that calculating the linguistic features over different contexts enables them to encode more information, whereas this is not the case for the WCFs.

The accuracies which we achieve with the slot and particle features calculated over all three contexts (an experiment which gives consistently high accuracies over verification and test data) correspond to a 30–31% reduction in error rate over the chance baseline for the 3-way task, and an 18–26% reduction in error rate for the 2-way task. Although we expected that the 2-way task may be easier, since it requires fewer fine-grained distinctions, it is clear that combining senses that have some motivation for being treated separately comes at a price.

The reductions in error rate that we achieve with our best features are quite respectable for a first attempt at addressing this problem, but more work clearly remains. In particular, there is a relatively high variability in performance across the verification and test sets, indicating that we need a larger number of experimental expressions to be able to draw firmer conclusions.

Chapter 7

Conclusions

While progress has recently been made in techniques for assessing the compositionality of a VPC, work thus far has left unaddressed the problem of determining the particular meaning of the component verb and particle. We have focused here on the semantic contribution of the particle—a part-of-speech whose semantic complexity and range of metaphorical meaning extensions has been largely overlooked in prior computational work. We adopted a cognitive linguistic perspective, and assumed that all VPCs are (at least partly) compositional and classified VPCs according to the meaning contributed by their particle. We developed features that capture linguistic properties of VPCs that are relevant to the semantics and syntax of particles and verbs, and showed that they outperform linguistically uninformed word co-occurrence features, achieving around a 20–30% reduction in error rate over a chance baseline.

7.1 Summary of Contributions

Particle Sense Classification While previous studies have focused on VPC compositionality, they have not addressed the issue of which meaning of the verb and particle is contributed. This study was the first to consider the issue of determining the semantics of the components of a VPC, and did so by classifying VPCs according to the meaning contributed by their particle.

Cognitive Linguistic Motivation Previous computational studies of VPCs have determined whether, or the extent to which, a VPC is compositional; doing so makes the assumption that some particles do not contribute any meaning to a VPC. In contrast, we based our classification of particle semantics on a cognitive linguistic analysis, which assumes that all VPCs are compositional. The network structure into which the senses of a particle may be organized allowed us to alter the granularity of classification in a linguistically motivated fashion.

Feature Space A set of features for particle sense classification was developed. Operating under the hypothesis that particle semantics is related to verb semantics, we demonstrated that features which have previously been used for the semantic classification of verbs can be directly used for the semantic classification of particles. A new set of features based on syntactic and semantic properties of particles was developed. Although these features were not found to perform well on their own, they did give more consistent performance across datasets when used in combination with other complementary linguistically motivated features.

Sense Annotation of Data We annotated a set of 180 VPCs using the particle *up* according to the sense class contributed by their particle. This data can be used in further studies of VPC semantics.

7.2 Limitations and Future Work

Size of Datasets The training, verification and test sets used in this study were each composed of sixty VPCs. Larger datasets were not used due to the expense of manually annotating the data. As discussed throughout Chapter 6, there was a large amount of variation between results on verification and test data, which may have been due to the small sample sizes used. Larger datasets would have allowed us to draw firmer conclusions from our results.

Type Classification In this study, we chose to focus on type-based, as opposed to token-based, classification. We did so partly because of the significant extra expense of manually annotating sufficient numbers of tokens in text. Previous studies into the semantic classification of verb tokens have shown information about the semantic class of a verb type to be a useful prior for a naive Bayes classifier (Lapata and Brew, 2004). We similarly believe that the semantic contribution of a particle in a VPC type is an informative prior for token-based particle sense classification. Thus our work will be a useful component of future work on the semantics of VPC tokens.

WCFs Features which are motivated by semantic and syntactic properties of verbs and VPCs seemed to outperform linguistically uninformed WCFs for the task of particle sense classification. However, the WCFs employed in this study were very simple. One modification of these features would be to restrict them to the arguments of a verb, as is noted in Section 6.4. This may allow them to encode more linguistically relevant information, and therefore perform better.

Additional Particles Although this study focused on the particle *up*, the feature space which we developed does not rely on syntactic and semantic properties which are specific to this particle. Therefore, our work may form the basis for future studies on the semantic classification of other particles.

Natural Language Understanding In order to effectively perform natural language processing tasks such as automatic machine translation and automatic text summarization, the semantics of each lexical item in a text must be determined. The coverage of

VPCs in lexical resources is known to be sparse, and furthermore, new combinations are constantly being generated. Therefore, automatic methods for determining the semantics of the components of a VPC are essential to improving the performance of natural language processing systems.

Appendix A

Human Judgements of Particle Sense Contribution

The following tables show the particle sense contribution judgements used in this study. The columns of the tables contain the following information:

Verb The VPC in question is composed of this verb and *up*.

Judge 1 The sense contributed by *up* according to judge 1.

Judge 2 The sense contributed by *up* according to judge 2.

Final The sense contributed by *up* used in classification tasks in this study. This corresponds to either the common sense given by both judges, or to the sense reached by the judges after discussion when their individual judgements disagreed.

VPCs on which the judges disagreed are shown in bold.

Verb	Judge 1	Judge 2	Final
balance	Cmpl	Cmpl	Cmpl
beat	Cmpl	Cmpl	Cmpl
bind	Refl	Refl	Refl
bolster	Vert	Vert	Vert
bottle	**Refl**	**Cmpl**	**Refl**
bugger	Cmpl	Cmpl	Cmpl
bunch	Refl	Refl	Refl
button	Refl	Refl	Refl
connect	Refl	Refl	Refl
couple	Refl	Refl	Refl
cuddle	Refl	Refl	Refl
dish	Vert	Vert	Vert
draw	**Goal**	**Vert**	**Vert**
drink	Cmpl	Cmpl	Cmpl
feel	Vert	Vert	Vert
firm	**Vert**	**Cmpl**	**Cmpl**
flare	Vert	Vert	Vert
freeze	Cmpl	Cmpl	Cmpl
give	Vert	Vert	Vert
go	Vert	Vert	Vert
grow	Vert	Vert	Vert
heat	Vert	Vert	Vert
keep	**Vert**	**Goal**	**Vert**
lace	Refl	Refl	Refl
let	Vert	Vert	Vert
lift	Vert	Vert	Vert
load	Cmpl	Cmpl	Cmpl
loosen	**Vert**	**Cmpl**	**Cmpl**
measure	Vert	Vert	Vert
move	**Goal**	**Vert**	**Goal**

Verb	Judge 2	Judge 2	Final
muck	Cmpl	Cmpl	Cmpl
notch	**Vert**	**Cmpl**	**Vert**
pair	Refl	Refl	Refl
pile	Vert	Vert	Vert
prick	Vert	Vert	Vert
quicken	Vert	Vert	Vert
rest	Cmpl	Cmpl	Cmpl
roll	Refl	Refl	Refl
scratch	Cmpl	Cmpl	Cmpl
screw	Cmpl	Cmpl	Cmpl
shut	**Refl**	**Cmpl**	**Refl**
slip	Cmpl	Cmpl	Cmpl
slow	Vert	Vert	Vert
squeeze	Refl	Refl	Refl
stay	Vert	Vert	Vert
steam	Vert	Vert	Vert
stick	Vert	Vert	Vert
store	Vert	Vert	Vert
strap	Refl	Refl	Refl
sum	**Vert**	**Cmpl**	**Cmpl**
take	Vert	Vert	Vert
tape	Refl	Refl	Refl
tone	**Vert**	**Cmpl**	**Cmpl**
train	**Vert**	**Cmpl**	**Cmpl**
twist	**Vert**	**Refl**	**Cmpl**
use	Cmpl	Cmpl	Cmpl
wax	Cmpl	Cmpl	Cmpl
whip	Vert	Vert	Vert
wipe	Cmpl	Cmpl	Cmpl
zip	Refl	Refl	Refl

Table A.1: Human annotator judgements for training set.

Verb	Judge 1	Judge 2	Final
act	**Vert**	**Cmpl**	**Cmpl**
blow	Cmpl	Cmpl	Cmpl
bring	Vert	Vert	Vert
brush	Cmpl	Cmpl	Cmpl
call	Vert	Vert	Vert
chalk	**Cmpl**	**Vert**	**Cmpl**
clear	Cmpl	Cmpl	Cmpl
clock	**Cmpl**	**Vert**	**Vert**
cough	Vert	Vert	Vert
cover	Cmpl	Cmpl	Cmpl
curl	Refl	Refl	Refl
drag	Vert	Vert	Vert
dream	Vert	Vert	Vert
dress	**Vert**	**Cmpl**	**Vert**
drum	Vert	Vert	Vert
fire	Vert	Vert	Vert
foul	Cmpl	Cmpl	Cmpl
fry	**Cmpl**	**Vert**	**Cmpl**
gear	Vert	Vert	Vert
get	Vert	Vert	Vert
gobble	Cmpl	Cmpl	Cmpl
head	Vert	Vert	Vert
hole	**Vert**	**Refl**	**Refl**
key	**Vert**	**Cmpl**	**Cmpl**
kiss	Goal	Goal	Goal
mock	Vert	Vert	Vert
nick	Cmpl	Cmpl	Cmpl
paste	Vert	Vert	Vert
prop	Vert	Vert	Vert
rake	Cmpl	Cmpl	Cmpl

Verb	Judge 1	Judge 2	Final
read	Cmpl	Cmpl	Cmpl
rustle	Vert	Vert	Vert
save	Vert	Vert	Vert
saw	Cmpl	Cmpl	Cmpl
send	Vert	Vert	Vert
serve	Vert	Vert	Vert
sober	Cmpl	Cmpl	Cmpl
soften	**Cmpl**	**Vert**	**Cmpl**
speak	Vert	Vert	Vert
speed	Vert	Vert	Vert
spoon	**Cmpl**	**Vert**	**Vert**
spring	Vert	Vert	Vert
staff	**Vert**	**Cmpl**	**Vert**
start	Vert	Vert	Vert
step	Vert	Vert	Vert
stir	Cmpl	Cmpl	Cmpl
straighten	**Vert**	**Cmpl**	**Vert**
study	Cmpl	Cmpl	Cmpl
stuff	Cmpl	Cmpl	Cmpl
surge	Vert	Vert	Vert
team	Refl	Refl	Refl
tear	Cmpl	Cmpl	Cmpl
think	Vert	Vert	Vert
tidy	Cmpl	Cmpl	Cmpl
toss	Vert	Vert	Vert
trade	Vert	Vert	Vert
tuck	**Vert**	**Refl**	**Vert**
weigh	Cmpl	Cmpl	Cmpl
work	Vert	Vert	Vert
wrap	Cmpl	Cmpl	Cmpl

Table A.2: Human annotator judgements for verification set.

Verb	Judge 1	Judge 2	Final
back	Goal	Goal	Goal
brew	Vert	Vert	Vert
bubble	Vert	Vert	Vert
buck	Vert	Vert	Vert
build	Vert	Vert	Vert
cart	Vert	Vert	Vert
chop	Cmpl	Cmpl	Cmpl
clean	Cmpl	Cmpl	Cmpl
coil	Refl	Refl	Refl
count	**Cmpl**	**Vert**	**Cmpl**
crease	Cmpl	Cmpl	Cmpl
crush	**Cmpl**	**Refl**	**Refl**
double	Refl	Refl	Refl
end	Cmpl	Cmpl	Cmpl
fill	Cmpl	Cmpl	Cmpl
finish	Cmpl	Cmpl	Cmpl
fold	Refl	Refl	Refl
follow	Vert	Vert	Vert
free	Vert	Vert	Vert
grab	Vert	Vert	Vert
heal	Cmpl	Cmpl	Cmpl
hitch	Refl	Refl	Refl
hold	**Vert**	**Refl**	**Refl**
hook	Refl	Refl	Refl
jumble	Cmpl	Cmpl	Cmpl
knock	Cmpl	Cmpl	Cmpl
lap	**Cmpl**	**Vert**	**Vert**
lead	**Goal**	**Vert**	**Goal**
light	Vert	Vert	Vert
lighten	**Cmpl**	**Vert**	**Vert**

Verb	Judge 1	Judge 2	Final
line	Cmpl	Cmpl	Cmpl
march	**Goal**	**Vert**	**Goal**
mark	Cmpl	Cmpl	Cmpl
mess	Cmpl	Cmpl	Cmpl
muddle	Cmpl	Cmpl	Cmpl
open	**Vert**	**Cmpl**	**Cmpl**
order	Vert	Vert	Vert
pack	**Cmpl**	**Refl**	**Refl**
pay	Vert	Vert	Vert
phone	Vert	Vert	Vert
puff	**Vert**	**Cmpl**	**Vert**
pump	**Vert**	**Cmpl**	**Vert**
queue	Cmpl	Cmpl	Cmpl
raise	Vert	Vert	Vert
rear	Vert	Vert	Vert
ring	Vert	Vert	Vert
scale	Vert	Vert	Vert
settle	Cmpl	Cmpl	Cmpl
show	Vert	Vert	Vert
size	Cmpl	Cmpl	Cmpl
stoke	Vert	Vert	Vert
stop	Cmpl	Cmpl	Cmpl
sweeten	Vert	Vert	Vert
swell	Vert	Vert	Vert
throw	Vert	Vert	Vert
tip	Vert	Vert	Vert
type	Cmpl	Cmpl	Cmpl
wait	**Vert**	**Goal**	**Vert**
wash	Cmpl	Cmpl	Cmpl
whisk	Cmpl	Cmpl	Cmpl

Table A.3: Human annotator judgements for test set.

Appendix B

Additional Experimental Results

	3-way Task		2-way Task	
Features	Ver	Test	Ver	Test
Slots	43	45	61	58
Slots + Particles	40	49	62	56
All Linguistic	43	47	60	57

Context: VPCs with *up* and verbs.

	3-way Task		2-way Task	
Features	Ver	Test	Ver	Test
Slots	44	51	47	58
Slots + Particles	48	37	58	60
All Linguistic	45	39	62	54

Context: VPCs with *up* and all VPCs.

Table B.1: Accuracy (%) using linguistic features.

	3-way Task		2-way Task	
Features	Ver	Test	Ver	Test
WCF_{200}	43	38	60	52
WCF_{500}	35	37	61	49

Context: VPCs with *up* and verbs.

	3-way Task		2-way Task	
Features	Ver	Test	Ver	Test
WCF_{200}	36	50	54	53
WCF_{500}	43	38	60	57

Context: VPCs with *up* and all VPCs.

Table B.2: Accuracy (%) using WCFs.

Features	3-way Task		2-way Task	
	Ver	Test	Ver	Test
Combined_{200}	48	47	65	56
Combined_{500}	65	47	62	49

Context: VPCs with *up* and verbs.

Features	3-way Task		2-way Task	
	Ver	Test	Ver	Test
Combined_{200}	42	42	63	54
Combined_{500}	56	42	67	47

Context: VPCs with *up* and all VPCs.

Table B.3: Accuracy (%) combining slot and particle features with WCFs. Note that the slot and particle features are calculated over all three contexts, while the WCFs are calculated for the specific contexts given above.

Features	3-way Task		2-way Task	
	Ver	Test	Ver	Test
Combined_{200}	56	51	65	59
Combined_{500}	57	47	59	52

Context: VPCs with *up* and verbs.

Features	3-way Task		2-way Task	
	Ver	Test	Ver	Test
Combined_{200}	45	45	62	49
Combined_{500}	56	45	59	52

Context: VPCs with *up* and all VPCs.

Table B.4: Accuracy (%) combining all linguistic features with WCFs. Note that the linguistic features are calculated over all three contexts, while the WCFs are calculated for the specific contexts given above.

Bibliography

S. Abney. 1991. Parsing by chunks. In R. Berwick, S. Abney, and C. Tenny, editors, *Principle-Based Parsing: Computation and Psycholinguistics*, p. 257–278. Kluwer Academic Publishers.

Y. S. Alam. 2004. Decision trees for sense disambiguation of prepositions: Case of over. In *HLT-NAACL 2004: Workshop on Computational Lexical Semantics*, p. 52–59.

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

T. Baldwin. 2005b. Looking for prepositional verbs in corpus data. In *Proceedings of the Second ACL-SIGSEM Workshop on the Linguistic Dimensions of Prepositions and their Use in Computational Linguistics Formalisms and Applications*, p. 115–126.

T. Baldwin. 2006. Distributional similarity and preposition semantics. In P. Saint-Dizier, editor, *Computational Linguistics Dimensions of Syntax and Semantics of Prepositions*, p. 197–210. Springer.

T. Baldwin and A. Villavicencio. 2002. Extracting the unextractable: A case study on verb-particles. In *Proceedings of the Sixth Workshop on Computational Language Learning (CoNLL-2002)*, p. 98–104.

C. Bannard. 2002. Statistical techniques for automatically inferring the semantics of verb-particle constructions. LinGO Working Paper No. 2002-06.

C. Bannard. 2005. Learning about the meaning of verb-particle constructions from corpora. *Computer Speech and Language, Special Issue on Multiword Expressions*, 19(4):467–478.

C. Bannard, T. Baldwin, and A. 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*, p. 65–72.

D. Biber, S. Johansson, G. Leech, S. Conrad, and E. Finegan. 1999. *Longman Grammar of Spoken and Written English.* Pearson Education Limited.

D. Blaheta and M. Johnson. 2001. Unsupervised learning of multi-word verbs. In *Proceedings of the 39th Annual Meeting of the Association for Computational Linguistics (ACL-2001)*, p. 54–60.

D. Bolinger. 1971. *The Phrasal Verb in English.* Harvard University Press.

C. Boonthum, S. Toida, and I. Levinstein. 2005. Sense disambiguation for preposition 'with'. In *Proceedings of the Second ACL-SIGSEM Workshop on the Linguistic Dimensions of Prepositions and their Use in Computational Linguistic Formalisms and Applications*, p. 153–162.

L. Burnard. 2000. *The British National Corpus Users Reference Guide.* Oxford University Computing Services.

B. Cappelle. 2002. And up it rises: particle preposing in English. In N. Dehe, R. Jackendoff, A. McIntyre, and S. Urban, editors, *Verb-Particle Explorations.* Mouton de Gruyter.

C.-C. Chang and C.-J. Lin. 2001. *LIBSVM: a library for support vector machines.* Software available at `http://www.csie.ntu.edu.tw/~cjlin/libsvm`.

S. C. Deerwester, S. T. Dumais, T. K. Landauer, G. W. Furnas, and R. A. Harshman. 1990. Indexing by latent semantic analysis. *Journal of the American Society of Information Science*, 41(6):391–407.

N. Dehé, R. Jackendoff, A. McIntyre, and S. Urban. 2002. *Verb-Particle Explorations.* Mouton de Gruyter.

B. Dorr. 2001. LCS verb database, online software database of lexical conceptual structures and documentation. Technical report, University of Maryland College Park.

A. Fazly, R. North, and S. Stevenson. 2005. Automatically distinguishing literal and figurative usages of highly polysemous verbs. In *Proceedings of the ACL-2005 Workshop on Deep Lexical Acquisition.*

C. Fellbaum, editor. 1998. *Wordnet: An Electronic Lexical Database.* Bradford Books.

C. J. Fillmore, C. Wooters, and C. F. Baker. 2001. Building a large lexical databank which provides deep semantics. In *Proceedings of the Pacific Asian Conference on Language, Information and Computation.*

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

S. T. Gries. 2002. The influence of processing on syntactic variation: Particle placement in English. In N. Dehe, R. Jackendoff, A. McIntyre, and S. Urban, editors, *Verb-Particle Explorations.* Mouton de Gruyter.

B. Hampe. 2000. Facing up to the meaning of 'face up to': A cognitive semantico-pragmatic analysis of an English verb-particle construction. In A. Foolen and F. van der Leek, editors, *Constructions in Cognitive Linguistics. Selected Papers from the fifth*

International Cognitive Linguistics Conference, p. 81–101. John Benjamins Publishing Company.

J. A. Hawkins. 1994. *A Performance Theory of Order and Constituency*. Cambridge University Press.

C.-W. Hsu, C.-C. Chang, and C.-J. Lin. 2003. *A Practical Guide to Support Vector Classification*. `http://www.csie.ntu.edu.tw/~cjlin/libsvm/`.

R. Jackendoff. 2002. English particle constructions, the lexicon, and the autonomy of syntax. In N. Dehe, R. Jackendoff, A. McIntyre, and S. Urban, editors, *Verb-Particle Explorations*. Mouton de Gruyter.

E. Joanis. 2002. Automatic verb classification using a general feature space. Master's thesis, Department of Computer Science, University of Toronto.

E. Joanis and S. Stevenson. 2003. A general feature space for automatic verb classification. In *Proceedings of the Conference of the European Chapter of the Association for Computational Linguistics (EACL-2003)*, p. 163–170.

E. Joanis, S. Stevenson, and D. James. 2006. A general feature space for automatic verb classification. Under journal review.

G. Lakoff and M. Johnson. 1980. *Metaphors We Live By*. University of Chicago Comment Press, Chicago.

R. W. Langacker. 1987. *Foundations of Cognitive Grammar: Theoretical Prerequisites*, volume 1. Stanford University Press, Stanford.

M. Lapata and C. Brew. 2004. Verb class disambiguation using informative priors. *Computational Linguistics*, 30(1):45–73.

B. Levin. 1993. *English Verb Classes and Alternations: A Preliminary Investigation*. University of Chicago Press, Chicago.

D. Lin. 1999. Automatic identification of non-compositional phrases. In *Proceedings of the 37th Annual Meeting of the Association for Computational Linguistics*, p. 317–324.

S. Lindner. 1981. *A lexico-semantic analysis of English verb particle constructions with out and up*. Ph.D. thesis, University of California, San Diego.

S. Lindner. 1982. What goes up doesn't necessarily come down: The ins and outs of opposites. In *Papers from the Eighteenth Regional Meeting, Chicago Linguistic Society*, p. 305–323.

K. C. Litkowski. 2005. The Preposition Project. In *Proceedings of the Second ACL-SIGSEM Workshop on the Linguistic Dimensions of Prepositions and their Use in Computational Linguistics Formalisms and Applications*.

K. C. Litkowski. 2002. Digraph analysis of dictionary preposition definitions. In *Proceedings of the SIGLEX/SENSEVAL Workshop on Word Sense Disambiguation: Recent Successes and Future Directions*, p. 9–16.

B. Lohse. 2004. Domain minimization in English verb-particle constructions. *Language*, 80(2):238–261.

M. Marcus, G. Kim, M. Marcinkiewicz, R. MacIntyre, A. Bies, M. Ferguson, K. Katz, and B. Schasberger. 1994. The Penn Treebank: Annotating predicate argument structure. In *Proceedings of the ARPA Human Language Technology Workshop*.

D. McCarthy, B. Keller, and J. Carroll. 2003. Detecting a continuum of compositionality in phrasal verbs. In *Proceedings of the ACL-SIGLEX Workshop on Multiword Expressions: Analysis, Acquisition and Treatment*.

D. McCarthy, R. Koeling, J. Weeds, and J. Carroll. 2004. Finding predominant word senses in untagged text. In *Proceedings of the 42nd Annual Meeting of the Association for Computational Linguistics*, p. 280–287.

A. McIntyre. 2001. *The particle verb list*. `http://www.uni-leipzig.de/~angling/mcintyre/pv.list.pdf`.

A. McIntyre. 2002. Idiosyncrasy in particle verbs. In N. Dehe, R. Jackendoff, A. McIntyre, and S. Urban, editors, *Verb-Particle Explorations*. Mouton de Gruyter.

I. D. Melamed. 1997. Automatic discovery of non-compositional compounds in parallel data. In *Proceedings of the Second Conference on Empirical Methods in Natural Language Processing*, p. 97–108.

P. Merlo and S. Stevenson. 2001. Automatic verb classification based on statistical distributions of argument structure. *Computational Linguistics*, 27(3):373–408.

P. S. Morgan. 1997. Figuring out *figure out*: Metaphor and the semantics of the English verb-particle construction. *Cognitive Linguistics*, 8(4):327–357.

T. O'Hara and J. Wiebe. 2003. Preposition semantic classification via Penn Treebank and FrameNet. In *Proceedings of the Seventh Conference on Natural Language Learning (CoNLL-2003)*, p. 79–86.

J. Patrick and J. Fletcher. 2005. Classifying verb-particle constructions by verb arguments. In *Proceedings of the Second ACL-SIGSEM Workshop on the Linguistic Dimensions of Prepositions and their use in Computational Linguistics Formalisms and Applications*, p. 200–209.

I. A. Sag, T. Baldwin, F. Bond, A. Copestake, and D. Flickinger. 2002. Multiword expressions: A pain in the neck for NLP. In *Proceedings of the Third International Conference on Intelligent Text Processing and Computational Linguistics (CICLING 2002)*, p. 1–15.

P. Saint-Dizier. 2005. PrepNet: a framework for describing prepositions: Preliminary investigation results. In *Proceedings of the Sixth International Workshop on Computational Semantics (IWCS'05)*, p. 145–157.

W. S. Sarle. 2002. Should I nonlinearly transform the data? In *Neural Network FAQ, part 2 of 7: Learning.* Periodic posting to the Usenet newsgroup `comp.ai.neural-nets`, URL: `ftp://ftp.sas.com/pub/neural/FAQ.html`.

S. Schulte im Walde. 2003. Experiments on the choice of features for learning verb classes. In *Proceedings of the 10th Conference of the European Chapter of the Association for Computational Linguistics (EACL-2003)*, p. 315–322.

S. Schulte im Walde. 2004. Identification, quantitative description, and preliminary distributional analysis of German particle verbs. In *Proceedings of the COLING Workshop on Enhancing and Using Electronic Dictionaries.*

S. Schulte im Walde. 2005. Exploring features to identify semantic nearest neighbours: A case study on German particle verbs. In *Proceedings of the International Conference on Recent Advances in Natural Language Processing.*

K. Uchiyama, T. Baldwin, and S. Ishizaki. 2005. Disambiguating Japanese compound verbs. *Computer Speech and Language, Special Issue on Multiword Expressions*, 19(4):497–512.

A. Villavicencio. 2003. Verb-particle constructions in the World Wide Web. In *Proceedings of the Second ACL-SIGSEM Workshop on the Linguistic Dimensions of Prepositions and their use in Computational Linguistics Formalisms and Applications.*

A. Villavicencio. 2005. The availability of verb-particle constructions in lexical resources: How much is enough? *Computer Speech and Language, Special Issue on Multiword Expressions*, 19(4):415–432.

A. Villavicencio, F. Bond, A. Korhonen, and D. McCarthy. 2005. Introduction to the special issue on multiword expressions: Having a crack at a hard nut. *Computer Speech and Language, Special Issue on Multiword Expressions*, 19(4):365–377.

S. Wurmbrand. 2000. The structure(s) of particle verbs. Master's thesis, McGill University.