

Context Resolution of Verb Particle Constructions for English to Hindi Translation

Niladri Chatterjee, Renu Balyan

Department of Mathematics, Indian Institute of Technology Delhi, India
{niladri.iitd, renu17775}@gmail.com

Abstract. Verb Particle Constructions (VPCs) are flexible in nature and hence quite complex and challenging to handle. As a consequence, VPCs generate a lot of interest for NLP community. Despite their prevalence in English they are not handled very well, and hence often result in poor quality of translation. In this paper we investigate VPCs for English to Hindi translation. An English VPC can have different meanings in Hindi based on what its neighboring entities are. The paper focuses on finding the correct Hindi verb for an English VPC. We also discuss some rules for VPC identification, and approaches for resolving the context of a VPC for English to Hindi machine translation.

Keywords: VPC, English to Hindi Machine Translation, Semantics, Separability.

1 Introduction

Quality of a translation often depends upon how correctly the verb of the source language sentence is translated into the target language. In a sentence the verb acts as the binding agent, and is therefore considered the most important component of a sentence. Consequently, verbs need to be handled in a more systematic way for a Machine Translation (MT) system. This is more pertinent with respect to translation from English to Indian languages, Hindi in particular. This is because English verbs are often polysemic, whereas Hindi has different verbs for each of the senses. For illustration, according to WordNet¹, the verb “break” has 59 senses, “make” has 49 senses, “give” has 44, “get” has 36 senses and so on. Almost for each of these senses a different and specific verb exists in Hindi. Consequently, selection of an appropriate Hindi verb is very important during translation from English to Hindi (or any other target language). This can be determined by identifying the sense in which the English verb is used. But it has been observed that this is not so straightforward, as often other features like semantics of other components of the sentence have to be looked into for correct translation.

English verbs can be classified as: “Single verbs” and “Phrasal verbs” (Singh, 2003). Single verbs are formed using only single words, e.g. “go”, “decide”, “take”, and “give”. Phrasal verbs are made of two or more words, like “take off”, “make up”, “do away with”, “put up with” etc. Our focus in this paper is on translation of this special category of verbs, the ‘phrasal verbs’ from English to Hindi. The motive here is to capture the correct sense and hence provide correct translation of phrasal verbs in Hindi, as their structures. In the rest of the paper we shall be using the term VPC instead of “phrasal verb”.

The difficulty in handling English VPCs is often due to their flexible syntactic structure in a sentence. Consider, for example, VPCs comprising a verb followed by a preposition. This can happen in two ways:

- a) Intransitive verbs are followed by some prepositions e.g. “refer to”, “look at”.
- b) Some VPCs have the structure of verb + preposition to convey some exact sense, e.g.

¹ <http://wordnet.princeton.edu> (Version 2.1)

“put out”, “put off”, and “run into”.

Hence the primary difficulty is to understand the sense of the *verb + preposition* combination. But the difficulty here is that in some cases the prepositions in phrasal verbs can appear before or after an object, whereas in other cases they can appear only before the object and never after it. Some examples to illustrate this point are:

(a) She *switched off* the light vs. She *switched* the light *off*.

(b) The customer *threw away* the plates vs. The customer *threw* the plates *away*.

In examples (a) and (b) prepositions “off” and “away” can appear before as well as after the object (light in example (a) and plates in example (b)). However, in examples (c) and (d) below the preposition “out” can appear only before the object (change and solution, respectively) and not after it. It is worth noting that the preposition “out” in both the sentences is part of the VPC, and not because the verbs concerned are intransitive.

(c) The new manager *brought out* a change.

(d) The programmer *found out* a solution.

VPCs are generally not handled properly in the existing English to Hindi Machine Translation (MT) systems. Table 1 shows translations of VPCs for two simple sentences as given by some of the most commonly used English to Hindi MT systems, namely Google², MANTRA³, MaTra⁴ and Anuvadakh⁵, which we shall refer to as MT1, MT2, MT3 and MT4, respectively in Table 1.

Table 1: Translation of VPCs by MT systems

Sentence	Hindi Translations	Problems encountered
He <i>switched on</i> the light.	MT1: <i>vah prakaash par bandh kar diyaa</i> He light on close did MT2: <i>vah ne roshni par badlaa</i> He light on changed MT3: <i>unhone prakaash par switch dabaayaa</i> By him light on switch pressed MT4: <i>unhone prakaash par badlaa kiyaa</i> By him light on change did	- Sense of “switch” is incorrect in MT1, MT2 and MT4. - The preposition “on” is taken as an entity independent of the verb “switch” by all the systems.
He <i>put on</i> the shoes.	MT1: <i>vah joota par daal diyaa</i> He shoes on spill did MT2: <i>vah joota daalte hein</i> He shoes wears MT3: <i>unhone jootain par daal</i> By him shoes on spill MT4: <i>unhone joota rakhtaa hai</i> By him shoes keeps	- “Put on” not treated as single unit by MT1 and MT3. - Incorrect Sense of “put” by all except MT2. - Incorrect tense by all the four systems.

This shows the need and necessity to look into these issues and thus design methods to handle these problems. In this work we focus on this aspect of English to Hindi MT in detail.

The paper is organized as follows. Section 2 discusses some of the previous works done for verb particle constructions (VPCs) and the problems faced during the English to Hindi translation of a VPC. A detailed analysis of the VPCs with their respective Hindi verbs based on the senses, and the VPC separability are also discussed in Section 2. Various rules for identification of a

² <http://translate.google.com/>

³ <http://mantra-rajbhasha.cdac.in/mantrarajbhasha/>

⁴ <http://www.cdacmumbai.in/matra/>

⁵ <http://tdil-dc.in/>

VPC and some methods for resolving the context for VPCs for English to Hindi machine translation are discussed in Section 3. Section 4 concludes the paper.

2 Verb Particle Constructions

Verb particle constructions (VPCs) consist of a head verb and one or more obligatory particles, in the form of intransitive prepositions (e.g. “hand in”, “take off”), adjectives (e.g. “cut short”) or verbs (e.g. “let go”, “let fly”). The focus of this paper is on prepositional particles due to their high frequency in text. VPCs cause significant problems for NLP systems. Semantically, they are seldom understood through the simple composition of their independent parts (Bannard *et al.*, 2003). The VPCs can display both a *joined* or *split* configuration. This prevents the systems from treating them as a single unit. Therefore, one prime requirement for an MT system is to be able to maintain the semantic bond between the components, even when they are non-adjacent (Constable and Curran, 2009).

The problem encountered for these prepositional VPCs has been shown for “put up” in the following sentences: (1) Ram *put* the picture *up*. vs. (2) Ram was *put up* in a hotel. When we compare, sentences (1) and (2), the meaning in (1) seems to be that Ram *put* the picture somewhere and that as a consequence the picture was *up*. The verb and the particle, therefore, make independent contributions to the sentence. Sentence (2), on the other hand requires a rather different analysis. Ram can neither be said to have been *put* or be *up* in the hotel.

The major problem therefore is to identify whether the first or the second kind of semantic representation is appropriate? Previous research on VPCs has focused on their automatic extraction and classification (Baldwin and Villavicencio, 2002; Villavicencio, 2003). However, systematic study about how they should be handled during translation for a particular language is noticeably lacking.

2.1 Figuring out VPCs-The History

A lot of work has been done on actually figuring out the VPCs. According to Ramisch *et al.* (2008) a three way classification is adopted by Deh’e (2002), where a VPC can be classified as compositional, idiomatic or aspectual, depending on its sense.

- In compositional VPCs the meaning of the construction is determined by the literal interpretations of the particle and the verb. These VPCs usually involve particles with directional or spatial meaning, and these can often be replaced by the appropriate directional PPs. For illustration in the sentence “She carried the bags in the house”, the particle “in” can be replaced with the directional preposition “into” without changing the sense of the sentence.
- Idiomatic VPCs, on the other hand, cannot have their meaning determined by interpreting their components literally (e.g. “get on”, meaning to be on friendly terms with someone).
- Aspectual VPCs, which have the particle providing the verb with an endpoint, suggesting that the action described by the verb is performed completely, thoroughly or continuously (e.g. “tear up” meaning to tear something into a lot of small pieces).

Below we present some characterization of VPCs as discussed by Bolinger (1971):

C1: In a transitive VPC the particle may come either before or after the NP (e.g. He *backed up* the team vs. He *backed* the team *up*). However, whether a particle can be separated or not from the verb depends on the degree of bonding between them, the size of the NP, and the kind of NP. In some cases this separability is incorrect. For illustration, of the two sentences “He *got off* the bus”, and “He *got* the bus *off*”, the latter one is incorrect.

C2: Unstressed personal pronouns must precede the particle. For illustration, “They *ate* it *up*” is correct, but “They *ate up* it” is not.

C3: If the particle precedes a simple definite NP, the particle does not take the NP as its object. For example, consider the sentence “He *brought along* his friend”. Here the particle “along” does not take the NP “his friend” as an object. Whereas in the sentence “He *slept in* the hotel” the particle “in” has a stronger bonding with the NP “the hotel”. Hence it is not acting as a VPC. In this paper we have considered the semantics of the entities succeeding the VPC, or appearing in between the verb and the particle for VPC identification.

2.2 Translation of VPCs in Hindi and their problems

When it comes to translating these VPCs it proves to be more difficult. One VPC may have multiple possible translations based on its neighboring entities and their related semantics. Let us consider the VPC “*put on*” and different sentences with their Hindi translations as shown in Table 2.

Table 2: Different translations for the VPC “*put on*”

	English Sentence	Hindi Translation	Hindi Verb for “ <i>Put on</i> ”
a.	He <i>put on</i> the dress	<i>usne dress pahanii</i>	<i>pahanaa</i> → “to wear”
b.	He <i>put on</i> the fan	<i>usne pankhaa chalaayaa</i>	<i>chalaanaa</i> → “to start”
c.	He <i>put on</i> the light	<i>usne light jalaayii</i>	<i>jalaanaa</i> → “to burn”
d.	He <i>put on</i> efforts	<i>usne mehanat kariii</i>	<i>karna</i> → “to do”
e.	She <i>put on</i> make-up	<i>usne make-up lagaayaa</i>	<i>lagaanaa</i> → “to apply”
f.	She <i>put on</i> weight	<i>uskaa vajan badh gayaa</i>	<i>badhaanaa</i> → “to increase”
g.	She <i>put on</i> the tea	<i>usne chaay banayii</i>	<i>banaanaa</i> → “to prepare”
h.	She <i>put</i> the book <i>on</i> the table	<i>usne mej par kitaab rakhii</i>	“Non-VPC”

All the English sentences in Table 2 (except (h)) are structurally same. Hence it is expected that the translation of these sentences will also be similar. The structure of all the translated sentences is also similar (except (f) which is a translation divergence (Gupta and Chatterjee, 2003)). However, the Hindi verbs being used in the translations are different and depend on the object as well as its semantics used. Thus finding the object and its semantics will play a very important role in deciding what the Hindi verb for a particular VPC will be.

We have identified 22 VPCs for the [present](#) study which according to WordNet have large number of senses. Some example English sentences and their corresponding Hindi translations have been considered for each of these senses. In certain cases the number of WordNet senses and the possible Hindi verbs for a VPC are exactly the same. For example the VPC “break down” as shown in Table 3.

Table 3: Analysis for the VPCs “break down”

WordNet Sense	Hindi verb	Separability	Related Verb exists	Semantics required
1, 5 (to analyze, to decompose)	<i>todnaa, baatnaa</i>	Yes	No	No
2 (make ineffective)	<i>samaapat karna</i>	Yes	No	No
3, 8 (lose control of one’s emotion, collapse)	<i>ronaa, tootnaa</i>	No	No	No
4 (seize to work)	<i>kharaab honaa</i>	No	No	No
6, 7 (cause to fall, collapse)	<i>giraanaa, todnaa</i>	Yes	No	No

It has however been found that generally the number of verbs possible in Hindi for a VPC is less as compared to the senses found in WordNet. Some cases have also been found where a Hindi verb exists for a VPC, but no corresponding sense is available in WordNet. For example “break into” as discussed in Table 4. The VPC “*break into*” has only 2 senses in WordNet; whereas when we use it for translating to Hindi, more than 2 cases have been found. It has also been observed that “break into” is a non-separable VPC. In case this occurs in a sentence with some NP in between the words “break” and “into” then it is actually not behaving as a VPC. Table 4 presents a detailed analysis of “break into” with respect to its translation into Hindi.

Table 4: Analysis for the VPC “break into”

WordNet Sense	Hindi verb	Separability	Related Verb exists	Semantics required
1 (to cry)	<i>ronaa</i>	No	Yes	No
2 (change pace)	<i>daudnaa</i>	No	Yes	No
- (to laugh, smile)	<i>hansnaa, muskuraanaa</i>	No	Yes	No
- (to enter forcibly)	<i>ghusnaa</i>	No	No	Yes

The cases shown in Table 4 can all be generalized using the rule that needs to check the NP, and extract the verb form related to it. For illustration,

- 1) Tears would relate to verb “cry” and hence can be translated to “*ronaa*”.
- 2) Laughter relates to “laugh” which is translated as “*hansnaa*” and so on.

However, for the last case discussed in Table 4 we cannot derive a verb directly for the NP following the VPC. Hence we need to find semantics of the object. For example, if the object is referring to a building/dwelling place then the corresponding Hindi verb would be “*ghusnaa*”.

Another category of VPCs comprises the ones for which we need the semantics of the object in the sentence to get the Hindi verbs. One such example is “put on”. The verb “put” has nine senses and so does the VPC “put on”. We have found seven Hindi verbs for the VPC “*put on*”. All these verbs do not actually match the senses found from WordNet. This has been shown in Table 5.

Table 5: Analysis for the “put on” VPC

WordNet Sense	Hindi verb	Separability	Related Verb exists	Semantics required
1 (to wear)	<i>pahanaa</i>	Yes	No	Yes
3 (to make/ prepare)	<i>banaanaa</i>	Yes	No	Yes
4 (to present)	<i>pesh/ prastut karna</i>	No	No	Yes
2, 5 (to add)	<i>jodnaa</i>	No	No	Yes
7 (to apply)	<i>lagaanaa</i>	No	No	Yes
9 (to get fat, increase weight)	<i>motaa honaa, vazan badhnaa</i>	No	No	Yes
6, 8 (mount, deceive)	-	-	-	-
- (to switch on, to start, to do)	<i>jalaanaa, chalaanaa, karna</i>	Yes, Yes, No	No	Yes

We could not find any corresponding Hindi verb for the senses 6 and 8 of “*put on*” in WordNet. Also three Hindi verbs (*jalaanaa*, *chalaanaa* and *karnaa*) have been found which do not have a relevant sense mentioned in WordNet. In these cases we find the semantics of the object and based on the semantics we use a relevant Hindi verb. This can be illustrated with the following example, “*Put on the light/fan/TV*”.

A similar analysis has been done for other VPCs which have two or more Hindi verbs corresponding to the senses obtained for those VPCs from WordNet. This will prove helpful in the translation process and improve the quality of the translation in a machine translation system.

2.3 Separability of the VPCs

No standard technique is available for judging separability of a VPC. Our scheme for ascertaining separability for a VPC is the following. We have marked a VPC as separable or non-separable based on its frequency of occurrence in the text. We have used four search engines (Google, Bing, AltaVista and Lycos/ HotBot) for finding the frequency of occurrence of the VPCs. Some of these have been shown in Table 6⁶.

Table 6: Frequency of occurrence of some VPCs

	Example VPC	Google	Bing	AltaVista	Lycos/ HotBot
(1)	a. <i>Bring out</i> a change	91,4000	38,40,00,000	754	274,000,000
	b. <i>Bring</i> a change <i>out</i>	5	7	22	9
(2)	a. <i>Break into</i> the house	6,750,000	26,70,00,000	129,000	192,000,000
	b. <i>Break</i> the house <i>into</i>	10,400	54	221	223
(3)	a. <i>Switch on</i> the light	16,900,000	13,10,00,000	142,000	129,000,000
	b. <i>Switch</i> the light <i>on</i>	1,930,000	11,00,00,000	166,000	108,000,000
(4)	a. <i>Put on</i> efforts	12,600	1,99,00,000	650	20,400,000
	b. <i>Put</i> efforts <i>on</i>	80,900	16,00,00,000	5470	115,000,000

Three cases have been found to exist for most of the VPCs:

- The frequency of occurrence of split configuration is very low as compared to that of the joined configuration. As shown for example (1) and (2) in Table 6. Our conclusion is: the VPC is non-separable and it is incorrect to use the split configuration in a sentence.
- The frequency of occurrence of both configurations is high but the frequency of joined configuration is higher than the split configuration. This has been illustrated by example (3) in Table 6. We conclude that the VPC is separable and both the split as well as joined configurations can be used.
- The frequency of occurrence of both configurations is high but the frequency of split configuration is higher than the joined configuration. This case is illustrated with example (4) in Table 6. Our conclusion is that the split configuration is not a VPC. Some other factors are contributing to its high frequency of occurrence. The joined configuration is a VPC. It is to be noted that this conclusion is based on the set of sentences considered for our experiments only. A stronger conclusion in this regard can be achieved only after experimenting with a large set of examples.

⁶ These experiments were carried out in August 2011.

3 VPCs for English to Hindi Translation

The first problem that an MT system needs to solve regarding translating VPCs is to decide whether the *verb + preposition* combination is a VPC. Since VPCs can occur in split as well as joined configurations in a sentence, it becomes mandatory to find when the verb and the particle are acting as a single unit, and when they are behaving as independent entities. Once this problem is solved, the correct Hindi verb can be found by considering the semantics of the object. The following methods have been suggested by different researchers:

- Lexical entry for expressions having fixed structure and meanings (Sinha and Thakur, 2004)
- Context - disambiguation rules (Sinha and Thakur, 2004; Bhandari *et al.*, 2002)
- Structural analysis of clauses (Sinha and Thakur, 2004)
- Use of Example-base approach for disambiguating the Phrasal verbs (Saha *et al.*, 2004)

In our approach the focus is more on the semantics of the object in a sentence. This helps in finding when a verb and a particle combination is a VPC and when it is not. Once it is known whether a VPC exists or not then the context of the VPC is resolved based on the same semantics. We first discuss certain rules based on which a VPC is identified and then describe various approaches that have been used for resolving the context of VPC if at all it exists.

3.1 Rules for VPC identification

The rules generated for identifying a VPC have been illustrated with some VPC examples, such as “put on”, “keep on”, “get off”, “go on” in this section. In the following discussion we use <verb name> to indicate that various inflections of the verb have been considered.

- Put on - The semantics considered for direct object (Obj₁) that immediately follow this VPC are: {Electric appliance: light, fan; wearable: dress, shoes; Eatable: tea, coffee; Chemical: cosmetics, paint; Public activity: play, dance; Quantity: weight, distance; Act: physical, mental activity}. The semantics considered for the indirect object (Obj₂), if at all it is present in the sentence are: {Location: wall, roof; Object: table, bed; Event: picnic, trip}. The various inflections of “put” that have been considered are “put”, “puts” and “putting”. The rules are as follows:
 - (a) IF <put> + on + Obj₁ (NP⁷₁). THEN a VPC.
 - (b) IF <put> + Obj₁ (NP₁) + on. THEN a VPC.
 - (c) IF <put> + Obj₁ (NP₁) + on + Obj₂ (NP₂). THEN not a VPC.
- Keep on - The semantic considered for object (Obj₁) is: {Thing: Living or Non- Living: fan, book, hat, dress, cosmetics, box, dog, cat etc.}. The semantics considered for the object (Obj₂), if at all it is present in a sentence are: {Location: wall, roof; Object: table, bed}. The semantic for the object (Obj₃) in rule (d) can only be some wearable item (e.g. hat, shoes, coat). The forms of the verb “keep” that have been used are: “keep” and “kept”. The rules are as follows:
 - (a) IF <keep> + on + [verb + ing] + ---. THEN a VPC.
 - (b) IF <keep> + Obj₁ (NP₁) + on + Obj₂ (NP₂). THEN not a VPC.
 - (c) IF <keep> + on + Obj₂ (NP₂). THEN not a VPC.
 - (d) IF <keep> + Obj₃ (NP) + on. THEN a VPC.
 - (e) IF <keep> + [verb + ing] + on + ---. THEN not a VPC.

⁷ Noun Phrase

- Get off - The semantics considered for object (Obj₁) are: {Vehicle: Bus, train; Riding animal: horse, elephant; Act: penalty, punishment}. The semantics considered for the object (Obj₂), are: {Attribute: Dirty mark, stain; Human being}. The various forms of get that have been used are: “get”, “got” and “getting”. The rules are as follows:
 - a) IF <get> + off + Obj₁ (NP₁) + ---. THEN a Non-separable VPC.
 - b) IF <get> + Obj₂ (NP₂) + off + ---. THEN a Separable VPC.
 - c) IF <get> + off + Obj₂ (NP₂) + ---. THEN a Separable VPC.

Similar rules have been formed for other VPCs that have been considered for study.

3.2 Context Resolution for VPCs translations

VPCs do not have a fixed behavior and exhibit different behavior in different contexts. The direct object in the sentence plays a key role in determining the sense of the VPC. The indirect object can also help in disambiguating the sense of the VPC (Saha *et al.*, 2004). In this section we suggest various methods for resolving the context of the VPCs for English to Hindi translation.

Lexical Entry

A number of VPCs (e.g. “go on”, “move on”, “call for”, “catch up”, “check in”) have a single Hindi verb for them. These VPCs are entered into the lexicon with their meaning. However, one needs to make sure in what situation they are acting as a VPC. The rules discussed in Section 3.1 determine when a VPC can occur in a sentence. If a VPC exists then it is given priority and the meaning of the VPC is picked from the lexicon instead of picking meanings of the individual components.

Sense of VPC in WordNet

WordNet has been used to decide the sense of the VPC in a sentence. Based on WordNet senses and the possible Hindi verbs available for these senses, rules have been formed to decide which Hindi verb has to be used for a VPC. We discuss here the rules for the VPC “break down” that have eight senses as per WordNet and can be translated using eight Hindi verbs. The task can also be fulfilled using only five unique Hindi verbs that are sufficient to represent all the eight senses. Some of these rules for the “break down” VPC are:

- IF (sense of “break down” is 1 or 5) THEN Hindi verb is “tohdanaa” or “baatnaa” meaning to break or to divide. For example, consider “I broke down the problem into sub-parts”.
- IF (sense of “break down” is 2) THEN Hindi verb is “samaapat karnaa” ~ to finish. For illustration, consider “The king tried to break down the racial discrimination”.

This approach is successful for only those VPCs for which all the senses have equivalent Hindi verbs and all the Hindi verbs can be represented by some sense in WordNet as in the case of “break down”. This approach fails for VPCs which have lesser number of senses in WordNet as compared to possible Hindi verbs, for instance the VPC “break into” and “put on”. For VPCs of this kind knowing only the sense from WordNet is not sufficient. Other possibilities also need to be explored. The following approach is useful in such a situation.

Related Verb for a Noun

This is applied when for an English VPC the number of senses in WordNet is not sufficient to indicate the Hindi verb to be used. For example “break into”. As discussed in Table 4, to deal with these cases one needs to check for the NP, if there is any, following the VPC. If the Noun in the NP has a related verb then translation of that verb is used to represent the Hindi for that VPC. For example, in the sentence “He broke into laughter/ tears” the VPC is followed by “laughter” which is associated with the verb “laugh” and hence the Hindi verb used is “hansnaa”. Similarly “tears” is associated with the verb “cry” and the Hindi verb for this is “ronaa”.

However, there are sentences, such as, “the thief *broke into* the house” in which we do not have a related verb for the NP (“the house”). In such a case the semantics of the object in the sentence is identified and then is used for context resolution. This method is explained below.

Semantics of the Object

If none of the approaches discussed above works, then we look for the object in the sentence and find its semantic. Based on the semantic of the object the Hindi verb that should be used to represent the VPC is found. Rules have been formed based on the semantics of the object. This has been illustrated for the VPC “put on” (see Table 5).

This VPC has nine senses in WordNet. Some senses (6 and 8) do not have corresponding Hindi verbs. Also three Hindi verbs were found where no corresponding sense occurs in WordNet. This is illustrated by the example “She *put on* the light”. No sense for the VPC exists in WordNet for this type of occurrence. Consequently, we found semantics of the object and decided the Hindi verb to be used accordingly. In total 18 such rules were formed for cases where no WordNet sense exists for a VPC but a Hindi verb is available. One of these rules is:

- IF (sense of “*put on*” is [NIL] and semantic of object is [Electric device: source of illumination or that emits light]) THEN Hindi verb will be “*jalaanaa*” ~ to burn

We studied 22 VPCs which had multiple Hindi verbs for their senses as found in WordNet. In all 166 senses were obtained for these VPCs from WordNet. However, most of these senses could be accommodated by 82 unique Hindi verbs. The synonyms of Hindi verbs have not been counted here. We also found 12 Hindi verbs which contributed to more than one WordNet sense obtained for different VPCs. One such Hindi verb is “*lagaanaa*”. This verb can be used for the VPC “put up” as well as “put on”. WordNet sense 1 for “put up” is “to post a sign or warning”; and WordNet sense 7 for “put on” is “to apply, something on a surface”. For illustration we consider the following two examples: (1) The college *put up* a notice for ragging. (2) She *put on* lipstick for the party. Both these senses can be represented by one Hindi verb “*lagaanaa*”. Thus we have lesser Hindi verbs to accommodate WordNet senses.

4 Concluding Remarks

This paper deals with handling of VPCs while translating from English to Hindi (E-H). Existing E-H MT systems are often found wanting while translating English VPCs. A thorough analysis of the outputs of four of the major E-H MT systems reveals the fact clearly. In particular, the problem becomes more prominent when the VPCs take the structure *verb + preposition*. The problem can be of various types. As a consequence the solution scheme needs to find answers to the following questions:

- whether the *verb + preposition* combination is a true VPC, as opposed to the preposition being used to take care of an intransitive verb;
- how to identify the VPC even when the preposition is not adjacent to the underlying verb;
- how to identify the exact sense of the VPC, given that the same VPC may have multiple senses, as per WordNet;
- how to choose the proper translation in the target language. This becomes very important in E-H translation, as for most of the English VPCs there exists a specific verb in Hindi.

None of the major E-H MT systems are able to handle these problems in an efficient way. We propose a rule-based approach for solving this difficulty. The rules are formulated by identifying the semantic bonding between the VPC and other components of the source language sentence. The rules have been formulated by analyzing a large collection of English sentences and their Hindi translations to identify the following:

- Different senses that a VPC may have.

- In which situations the verb and preposition can have a joined structure, and in which cases a split structure is also possible.
- In which situations the *verb + preposition* is not a VPC.

The rules have been applied for different VPCs with multiple senses, and the results are promising. So far we have considered only *Verb + Preposition* type of VPCs. However, English VPCs may take other forms too. Some of the common VPC structures are *Verb + Adjective* (e.g. “cut short”), *Verb + Preposition + Preposition* (e.g. “look down upon”). We would like to extend our scheme to handle these VPCs as well. Also, we have considered only simple sentences so far. Our intuitive feeling is that a study for complex sentences for these VPCs may lead to further refinement of the rules that have been discussed for identification and context resolution of the VPCs. We are currently working in these directions.

The approaches described are not language dependent. We expect that similar techniques will prove to be helpful for other target languages by focusing on the lexicon and the verbs of that particular language.

References

- Baldwin Timothy, Aline Villavicencio. 2002. Extracting the Unextractable: A Case Study on Verb-particles. *Proceedings of the 6th Conference on Natural Language Learning (CoNLL-2002)*, pp. 98-104. Taipei, Taiwan.
- Bannard Colin, Timothy Baldwin, Alex Lascarides. 2003. A Statistical Approach to the Semantics of Verb-Particles. *Proceedings of the ACL-SIGLEX Workshop on Multiword Expressions: Analysis, Acquisition and Treatment*, Sapporo, Japan, pp. 65-72.
- Bhandari Vartika, R.M.K. Sinha and Ajai Jain. 2002. Disambiguation of Phrasal Verbs for Machine Translation. *Proceedings of Symposium on Translation Support Systems (STRANS-2002)*, 15-17 March, Kanpur, India.
- Bolinger Dwight L. 1971. *The Phrasal Verb in English*. Boston, USA: Harvard University Press.
- Constable James W. D. and James R. Curran. 2009. Integrating verb-particle constructions into CCG parsing. *Proceedings of the Australasian Language Technology Association Workshop (ALTW)*, pp. 114–118.
- Deh'e, Nicole. 2002. *Particle verbs in English: syntax, information structure and intonation*. John Benjamins, Amsterdam/Philadelphia.
- Gupta Deepa and N. Chatterjee. 2003. Identification of Divergence for English to Hindi EBMT. *Proceedings of MT Summit IX*, New Orleans, LA, pp. 141 – 148.
- Ramisch Carlos, Aline Villavicencio, Leonardo Moura, Marco Idiart. 2008. Picking them up and Figuring them out: Verb-Particle Constructions, Noise and Idiomaticity. *Proceedings of the Twelfth Conference on Computational Natural Language Learning*, pp. 49-56.
- Saha Indranil, Ananthkrishnan R. and Sasikumar M. 2004. *Example-Based Technique for Disambiguating Phrasal Verbs in English to Hindi Translation*. Technical Report, KBCS Division, CDAC Mumbai.
- Singh Suraj Bhan. 2003. *English-Hindi Translation Grammar (A Transfer Grammar based on Contrastive Analysis)*. Prabhat Publications, New Delhi.
- Sinha R.M.K. and Anil Thakur. 2004. Multi-Word Expression in English and Hindi: Problems in Contextualization. *Proceedings of International Symposium on Machine Translation NLP and TSS (iSTRANS-2004)*, Vol. 1. pp. 111-116.
- Villavicencio Aline. 2003. Verb-Particle Constructions and Lexical Resources. *Proceedings of the ACL-SIGLEX Workshop on Multiword Expressions: Analysis, Acquisition and Treatment*, Sapporo, Japan.