

Chapter 1

Thesis Outlines

1.1 Research Topic

English phrasal verbs (e.g. pick up, give in, bring out, etc.) are well known for many reasons including for their structural and semantic complexities, challenges in pedagogical and computational processing applications. They have also been put to extensive studies in both theoretical and applied linguistics including corpus-based study as well as comparative and contrastive study to examine their mapping patterns in other languages. In the area of machine translation from English to other languages, in particular, the English phrasal verbs have been examined from the point to view of translation divergence (Dorr, 1993) in the sense that “translation divergence arises when the natural translation of one language into other results in a very different form than that of the original” (p. 19). The topic of translation divergence has received further attention in linguistics and grammar studies from different perspectives, particularly in comparative and contrastive study. In the recent decades, the topic has received attention from the point of view of the cross-linguistic mapping strategy of various lexical or phrasal expressions from one language to another. A lexical or phrasal expression in a language may have equivalent expressions in other languages with different morpho-syntactic structures and semantic properties. Complex cross-language relations between English lexical or phraseological expressions, or multiword expressions (MWE), and their translation equivalents in other languages can be clearly observed, highlighting their

relevance in diverse areas including language pedagogy, translation, automatic extraction, etc.

In the present study, English phrasal verbs (henceforth EPVs) are examined from the point of view of their mapping patterns and translation divergence in the context of obtaining their Hindi equivalents. As the grammatical systems of a pair of languages have different mechanisms to realise linguistic concepts, and also the more the two languages are distantly related (or not related at all), the higher the degree of divergence, our obvious and initial hypothesis is that in the mapping of the EPVs onto Hindi a higher degree of divergences are to be expected. A comparative study becomes quite necessary in the identification of divergences between two or more languages towards understanding the commonalities and the differences in their respective linguistic systems. Comparative (and contrastive) studies have been reported to serve several objectives including resources for foreign language teaching-learning and for natural language processing applications such as machine translation (Halliday, 1966; Dorr, 1993, 1994; Nerbonne, 2003). This has particularly close connection with foreign/second language teaching where it helps in understanding the commonalities and the differences in learner's first and the target languages including insights for the learner's errors and language interference. The importance of the comparative linguistic study easily extended to the area of machine translation (particularly in its initial rule-based stage) for obtaining mapping/transfer rules. Beginning with Dorr (1993, 1994), several pairs of languages taking English as the source language have been studied from this perspective (Chapter 3 presents a brief survey of some of these works). The EPVs have also attracted a good amount of attention, including in the research on multi-word expressions (MWEs), particularly for their complex semantic interpretations and compositionality (Tu & Roth, 2012). Some of the Indian languages including Hindi have also been compared and

contrasted with English to identify divergences (Dave et al., 2001; Sinha & Thakur, 2004, 2005a, 2005b, 2008; Goyal et al., 2004; Gupta & Chatterjee, 2003; Chatterjee & Balyan, 2011; among others). There also exists an initial comparative study of English and Hindi for the purpose of obtaining rules for mapping English phrasal verbs to Hindi (Chatterjee & Balyan, 2011). Chatterjee and Balyan (2011) proposed a semantic based disambiguation strategy to resolve the context of English phrasal verbs for obtaining the correct Hindi equivalent verbs for particular phrasal verbs in English for English-Hindi machine translation systems. They focused primarily on contextualising the phrasal verbs at the sentence and structural levels. It was neither their objective nor did they carry out a comprehensively detailed study to include the Hindi verbal systems and how they account for the semantics of English phrasal verb constructions. The present study is an attempt in this direction. However, the present study is not directly concerned with the identification and categorization task for the EPVs. For this, it uses the insights and findings obtained from some of the major corpus-based studies of EPVs (as detailed in the next chapter 2).

The corpus-based approach is one of the best suited options to carry out a study of this kind. Particularly, an English-Hindi parallel corpus would have made it less of a struggle in obtaining the necessary examples of the equivalent Hindi translations of the English phrasal verbs. However, such a corpus is not yet available (at least in the public domain). We made an initial attempt to use *samanantar* (Ramesh et al., 2022). However, but it was quickly observed that the quality of translations was not upto the mark and the corpus was full of inaccuracies, and we decided against using it further. This lack of necessary resources for the work made is imperative for us to develop EPVs related translation corpus for English-Hindi and because of this the work in the thesis has two important stages and components:

- i. creation of a translation corpus for selected EPVs-to-Hindi equivalents (in the form of an example base for the present study as prerequisite), and
- ii. the analysis of the data (from the corpus developed for this purpose) to study the mapping patterns of the English phrasal verbs in their Hindi equivalents.

The research questions (as listed below) are primarily centered around the second component of the work (an in ii above).

- i. How in Hindi the senses of the EPVs are mapped?

- that is, what are the possible verbal systems in Hindi to map the various senses of the EPVs?
- ii. How does Hindi represent the compositionally interpretable semantics (e.g. spatial directional and aspectual completive senses) of the respective particles in EPVs?
- iii. What kind of generalizations can be drawn from the findings of the study of the selected EPVs for rest of the EPVs?

The chapter 5 reports the works that have been attempted towards responding to these research questions.

1.2 Research Motivations

The English phrasal verbs (EPVs) are highly polysemous as they have distinct meanings in different constructions and contexts. This makes them quite challenging to obtain their translation in other languages, as they represent specific lexemes which typically cannot be transferred directly. Moreover, English phrasal verbs denote a range of meanings, from literal to idiomatic, which also causes difficulties in translation. Hindi, like many other

languages, does not have English-like phrasal verbs. Instead, Hindi has its own complex verbal system with several possible options to represent the various sense(s) of an English phrasal verb. A comparison of the EPVs with their Hindi equivalents to examine how in the Hindi verbal system the various senses of the English phrasal verbs including the semantics of the particle elements are mapped, presents an interesting research possibility and motivation from both linguistics and applied linguistics points of view. Hindi has compound and complex verbs besides simple lexical and other complex verbal constructions. The particle elements in the English phrasal verbs are observed to have complex semantic extensions and in Hindi they are represented by multiple possible equivalents. It has proved interesting to study this aspect of the EPVs from the cognitive linguistics perspective (reported in the second section of chapter 5).

As far as the applicational significance and relevance of this study are concerned, a study of this kind has diverse potential applications including language pedagogy and (machine) translation to lexicography and grammar. The study will also help to get insights for other Indian languages in the similar contexts. A deeper insight into the similarities and variations between English phrasal verbs and their Hindi counterparts holds significant importance for language learning-teaching. This has potential to assist learners in comprehending the subtleties of phrasal verbs in a manner that aligns with their current linguistic knowledge. From an educational standpoint, the findings from this study will contribute to an improved teaching-learning experience. For Hindi-speaking learners of English, examining these structures alongside their native language equivalents can create a connection between what they know and the new concepts they are encountering. This study provides insight into the ways various languages express similar concepts using unique structures. Examining these mappings uncovers a profound understanding of the syntax, semantics, and pragmatics inherent in both languages.

Further, this kind of comparative analysis is imperative for enhancing machine translation, language models, and various tools in the field of natural language processing. Phrasal verbs frequently present difficulties in translation systems because of their idiomatic meanings and context-specific interpretations. By identifying patterns and equivalences between English phrasal verbs and their Hindi counterparts, this study can contribute to the advancement of more precise translation algorithms. Furthermore, these investigations can enrich bilingual lexicons, advance sentiment analysis, and enhance text-generation models by providing a deeper understanding of idiomatic expressions in various languages. Therefore, we can say that a comparative study of English phrasal verbs and their Hindi equivalents holds considerable importance for language education, linguistic knowledge, and in the area of natural language processing.

1.3 Chapter Outlines

The thesis is organized in 6 chapters including the present one which presents brief outlines of the work. Chapter 2 reports some of the major corpus-based studies on the English phrasal verbs towards their identification, classification and sense determination on the basis of their frequency of occurrences and the number of senses they represent. Chapter 3 presents a brief literature survey of the comparative studies on the EPVs in selected pairs of languages. The chapter also presents a brief outline of the observations and insights obtained from these studies which have consequences for the study in the thesis. Chapter 4 attempts to outline the methods and resources that have been developed, accessed and used for carrying out the present study. The chapter includes a brief outline of the resources (corpus, dictionaries, etc) and also the theoretical assumptions (corpus-based comparative study and cognitive linguistics approach). Chapter 5 reports the main

work of the thesis. The chapter is divided into two main sections. The section one discusses and describes the EPVs and their possible Hindi equivalents, attempts to make a categorization based on their mapping patterns and reports the complexities in doing so. The section two examines how the compositional senses, primarily represented by the particle elements in the selected EPVs are mapped in the Hindi verbal system. Chapter 6 summarises the main points of the work.

