Handling of Prepositions in English to Bengali
Machine Translation

Sudip Kumar Naskar
Dept. of Comp. Sc. & Engg.,
Jadavpur University,
Kolkata, India
sudip_naskar@hotmail.com

Sivaji Bandyopadhyay
Dept. of Comp. Sc. & Engg.,
Jadavpur University,
Kolkata, India
sivaji_cse_ju@yahoo.com

Abstract

The present study focuses on the lexical meanings of prepositions rather than on the thematic meanings because it is intended for use in an English-Bengali machine translation (MT) system, where the meaning of a lexical unit must be preserved in the target language, even though it may take a different syntactic form in the source and target languages. Bengali is the fifth language in the world in terms of the number of native speakers and is an important language in India. There is no concept of preposition in Bengali. English prepositions are translated to Bengali by attaching appropriate inflections to the head noun of the prepositional phrase (PP), i.e., the object of the preposition. The choice of the inflection depends on the spelling pattern of the translated Bengali head noun. Further postpositional words may also appear in the Bengali translation for some prepositions. The choice of the appropriate postpositional word depends on the *WordNet* synset information of the head noun. Idiomatic or metaphoric PPs are translated into Bengali by looking into a bilingual example base. The analysis presented here is general and applicable for translation from English to many other Indo-Aryan languages that handle prepositions using inflections and postpositions.

1 Introduction

Prepositions have been studied from a variety of perspectives. Both linguistic and computational (monolingual and cross-lingual) aspects of prepositions have been contemplated by several researchers. Jackendoff (1977), Emonds (1985), Rauh (1993) and Pullum and Huddleston (2002) have investigated the syntactic characteristics of preposition. Cognitive theorists have examined the polysemous nature of prepositions and explored the conceptual relationships of the polysemy, proposing the graphical mental images (Lakoff and Johnson, 1980; Brugman, 1981, 1988; Herskovits, 1986; Langacker, 1987; Tyler and Evans, 2003). Fauconnier (1994) and Visetti and Cadiot (2002) have canvassed the pragmatic aspects of prepositions. A practical study of the usage of prepositions was carried out for the purpose of teaching English as a second language (Wahlen, 1995; Lindstromberg, 1997; Yates, 1999). The deictic properties of spatial prepositions have been studied by Hill (1982), while the geographical information provided by them was an interest of computational research (Xu and Badler, 2000; Tezuka et al., 2001).

In the fields of natural language processing, the problem of PP attachment has been a topic for research for quite a long time, and in recent years, the problem was explored with a neural network-based approach (Sopena et al., 1998) and with a syntax-based trainable approach (Yeh and Vilain, 1998). Although past research has revealed various aspects of prepositions, there is not much semantic research of prepositions available for computational use, which requires a vigorous formalization of representing the semantics. A recent semantic study of prepositions for computational use is found in (Voss, 2002), with a focus on spatial prepositions. Spatial prepositions are divided into three categories according to which one of the two thematic meanings between place and path they acquire when they are in argument, adjunct and non-subcategorized positions of particular types of

verbs. The semantics of spatial prepositions dealt with in (Voss, 2002) is not lexical but thematic.

There are some prepositions (e.g., over, with), which have many senses as preposition. By making use of the semantic features of the Complements (reference object) and Heads (verb, verb phrase, noun or noun phrase governing a preposition or a PP), the meaning of the polysemous prepositions can be computationally disambiguated. The different meanings of *over* call for different semantic features in its heads and complements [Alam, 04].

Prepositional systems across languages vary to a considerably degree, and this cross-linguistic diversity increases as we move from core, physical senses of prepositions into the metaphoric extensions of prepositional meaning (metaphor or rather, idiomaticity is one of the main realms of usage with prepositions) (Brala, 2000).

The present study focuses on the lexical meanings of prepositions rather than on the thematic meanings because it is intended for use in an English-Bengali machine translation (MT) system, where the meaning of a sentence, a phrase or a lexical entry of the source language must be preserved in the target language, even though it may take a different syntactic form in the source and target languages. Bengali is the fifth language in the world in terms of the number of native speakers and is an important language in India. It is the official language of neighboring Bangladesh. There is no concept of preposition in Bengali. English prepositions are translated to Bengali by attaching appropriate inflections to the head noun of the PP, i.e., the object of the preposition. The choice of the inflection depends on the spelling pattern of the translated Bengali head noun. Further postpositional words may also appear in the Bengali translation for some prepositions. The choice of the appropriate postpositional word depends on the *WordNet* (Fellbaum, 1998) synset information of the head noun. Idiomatic or metaphoric PPs are translated into Bengali by looking into a bilingual example base.

A brief overview of the English-Bengali MT System is presented in Section 2. Different types of English prepositions and their identification in the MT system are described in Section 3. Inflections and postpositions in Bengali are outlined in Section 4. Translation of English prepositions to inflections and postpositions in Bengali are detailed in Section 5. The conclusion is drawn in Section 6.

2 A Brief Overview of the English-Bengali MT System

The handling of English prepositions during translation to Bengali has been studied with respect to an English-Bengali MT system (Naskar and Bandyopadhyay, 2005) being developed. In order to translate from English to Bengali, the first thing we do is lexical analysis of the English sentence using the *WordNet*, to gather the lexical features of the morphemes. During morphological analysis, the root words / terms (including idioms and named entities), along with associated grammatical information and semantic categories are extracted. A shallow parser identifies the constituent phrases of the source language sentence and tags them to encode all relevant information that might be needed to translate these phrases and perhaps resolve ambiguities in other phrases. Then these phrases are translated individually to the target language (Bengali) using Bengali synthesis rules. The noun phrases and PPs are translated using Example bases of syntactic transfer rules. Verb phrase translation scheme is `rule based` and uses Morphological Paradigm Suffix Tables. Finally, those target language phrases are arranged using some heuristics, based on the word ordering rules of Bengali, to form the target language representation of the source language sentence.

3 Prepositions in English

A preposition is a word placed before a "noun" to show in what relation the noun stands with regard to the other noun and verb words in the same sentence. The noun that follows a preposition, i.e., the reference object is in the accusative case and is governed by the preposition. Prepositions can also be defined as words that begin *prepositional phrases* (PP). A PP is a group of words containing a *preposition*, an *object of the preposition*, and any *modifiers of the object*.

Syntactically, prepositions can be arranged into three classes – simple prepositions (e.g., *at, by, for, from* etc.), compound prepositions and phrase prepositions. A compound preposition is made up of a set of words which starts with and acts like a preposition (e.g., *in spite of, in favor of, on behalf of* etc.). A phrase preposition is a simple preposition preceded by a word from another category, such as an adverb, adjective, or conjunction (e.g., *instead of, prior to, because of, according to* etc.).

Frequently prepositions follow the verbs together forming phrasal verbs and remain sepa-

rate. A word that looks like a preposition but is actually part of a phrasal verb is often called a *particle*. E.g. "Four men held *up* the bank." Here *held up* is a verb ["to rob"]. Therefore, *up* is not a preposition, and *bank* is not the object of a preposition. Instead, bank is a direct object of the verb *held up*. A particle may not always appear immediately after the verb with which it makes up a phrasal verb (e.g., Four men held the bank *up*.).

An idiomatic (metaphoric) PP starts with a preposition, but its meaning cannot be ascertained from the meaning of its components. Examples of idiomatic PPs are: *at times*, *by hook or crook* etc.

All these syntactical characteristics are used to identify prepositions in the English-Bengali MT system. Moreover, the inventory of prepositions in English is a close set. So, identification of prepositions is not much of a problem in English. A simple list serves the purpose. The prepositions, compound prepositions, phrase prepositions and idiomatic PPs are identified during morphological analysis. Some of the phrasal verbs (when the phrasal verb appears as a whole) are identified during the morphological analysis phase and some during parsing (when the particle does not accompany the verb).

However, there are some words that act as prepositions and fall into other POS categories as well. For example, the word *before* can be used as an adverb (e.g., I could not come *before*), preposition (e.g., He came *before* me) or a conjunction (e.g., He came *before* I came). Similarly, the word *round* can be used as an adjective (e.g., Rugby is not played with a *round* ball), noun (e.g., Rafter was knocked out of the tournament in the third *round*), adverb (e.g., They have moved all the furniture *round*), preposition (e.g., The earth revolves *round* the sun) and verb (e.g., His eyes rounded with anger). But depending on the POS of the neighboring words/terms, the parser easily identifies the correct POS of the word in the particular context.

A preposition is usually placed in front of (is "pre-positioned" before) its object, but sometimes however may follow it (e.g., **What** are you looking *at*?). The preposition is often placed at the end when the reference object is an interrogative pronoun (e.g., **Where** are you coming *from*?) or a relative pronoun (e.g., My grandfather was a collector of coins, **which** we used to fight *over*). In such cases, the system finds out that the preposition is not a particle and is not followed by a noun either, so it must be a

stranded preposition. It searches the pronoun (relative or interrogative) that appears at its left and relates the stranded preposition to the pronoun. Thus during translation, the following conversion takes place.

```
(1)  Where  are  you  coming
from? ←→ From where are you
coming?
(2)  My  grandfather  was  a
collector of coins, which we
used to fight over. ←→ My
grandfather was a collector
of coins, over which we used
to fight.
```

But if the pronoun is missing, then the system has to find out the elliptical pronoun first.

```
(3) I am grateful to the man
I  have  spoken  to.  →  I  am
grateful to the man [whom] I
have  spoken  to.  →  I  am
grateful  to  the  man  to
[whom] I have spoken.
```

Prepositions represent several relations with the nouns governed by them. Spatial and temporal prepositions (which indicate a place or time relation) have received a relatively in-depth study for a number of languages. The semantics of other types of prepositions describing manner, instrument, amount or accompaniment largely remain unexplored. In case of an MT system, when a preposition has different representations in the target language for different relations indicated by it, identification of the relation is necessary. The *WordNet* synset information of the head noun of the PP, i.e., the object of the preposition serves to identify the relation.

4 Inflections and Postpositions in Bengali

In Bengali, there is no concept of preposition. English prepositions are handled in Bengali using inflections (*vibhakti*s) to the reference objects and/or post-positional words after them. Inflections get attached to the reference objects. An inflection has no existence of its own in the language, and it does not have any meaning as well. There are only a few inflections in Bengali: Φ (null), -কে(-*e*), -য় (-*y*), -য়ে (-*ye*), -তে (-*te*), -তেতে (-*ete*), -কে (-*ke*), -রে (-*re*), -কেরে (-*ere*), -র (-*r*) and -কের (-*er*) (an inflection is repre-

sented as a word with a leading '-' in this paper). The placeholder indicated by a dashed circle represents a consonant or a conjunct. For example, if -ে inflection is attached to the word বাজার (*bazar* [market]) the inflected word is বাজারে (*bazar-e* [market-to]). On the other hand, post-positional words are independent words. They have meanings of their own and are used independently like other words. A post-positional word is positioned after an inflected noun (the reference object). Some examples of the post-positional words in (colloquial) Bengali are: দিয়ে (*diye* [by]), থেকে (*theke* [from]), জন্য (*jonno* [for]), কাছে (*kachhe* [near]), সামনে (*samne* [in front of]) etc.

5 Translating English prepositions to Bengali

When an English PP is translated into Bengali, the following transformation takes place: (preposition) (reference object) ←→ (reference object) [(inflection)] [(postpositional-word)]. The correspondence between English prepositions and Bengali postpositions (inflections and postpositional words) is not direct. As far as the selection of the appropriate target language representation of a preposition is concerned the reference object plays a major role in determining the correct preposition sense. Deciding whether the preposition is used in a spatial sense, as opposed to a temporal or other senses, is determined by the semantics of the head noun of the reference object. A noun phrase (NP) denoting a place gives rise to a spatial PP. Similarly, an object referring to a time entity produces a temporal expression. These relationships can be established by looking at the *WordNet* synset information of the head noun of the PP.

5.1 Translating English prepositions using Inflections in Bengali

The translation of the three English prepositions 'in', 'on', and 'at' involves identifying the possible inflection to be attached to the head noun of the PP. No postpositional words are placed after the head noun for these prepositions. The three prepositions 'in', 'on', and 'at' (which are both spatial and temporal in nature) can be translated into the Bengali inflections '-ে' (*-e*), '-তে' (*-te*), -েতে (*-ete*) and 'য়' (*-y*). Any of these 4 Bengali inflections can be placed after the reference ob-

ject for any of these 3 English spatial and temporal prepositions. The choice depends on the spelling of the translated reference object. The rule is: if the last letter of the Bengali representation of the reference object is a consonant, 'ে' (*-e*) or -েতে (*-ete*) is added to it (e.g., at/in market→ বাজারে [*bazar-e* / *bazar-ete*]), else if the last letter of the Bengali word is a *matra* (vowel modifier) and if the *matra* is 'া' (*-a*), any of 'তে' (*-te*), or 'য়' (*-y*) can be added to the Bengali reference word (e.g., in evening→ সন্ধ্যাতে / সন্ধ্যায় [*sandhya-te* / *sandhya-y*]), otherwise 'তে' (*-te*) is added to it (e.g., at home→ বাড়িতে [*badi-te*]).

When translating the temporal expressions, if '*on*' is followed by a day (like Sunday, Monday etc.) or by a date in English, null inflection is added.

To translate this type of PPs, we take the help of an example base, which contains bilingual translation examples. Here are some translation examples from the example base (TLR – target language representation of the reference object).

```
(1)  at  /  in  (place)  ←→
(TLR) – (ে / য়ে / তে ) [ - ( e /
ye / te )]
(2) of (NP) ←→ (TLR) – ( র /
ের / য়ের ) [ – ( r / er / yer
)]
```

5.2 Translating English prepositions using Inflections and Postpositions in Bengali

Most of the English prepositions are translated to Bengali as inflections and postpositions to the noun word representing the reference object. To translate this type of PPs, we take the help of an example base, which contains bilingual translation examples. Here are some translation examples from the example base (TLR – target language representation of the reference object).

```
(1)  before  (artifact)  ←→
(TLR) – ( র / ের / য়ের ) সামনে [ –
( r / er / yer ) samne ]
(2)  before  (!artifact)  ←→
(TLR) – ( র / ের / য়ের ) আগে [ –
( r / er / yer ) age ]
```

```
(3) round (place / physical
object) ←→ (TLR) - ( র / ের /
যের ) চারদিকে [ - ( r / er / yer
) chardike ]
(4) after (time) ←→ (TLR) -
( র / ের / যের ) পরে [ - ( r / er
/ yer ) pare ]
(5) since (place / physical
object / time) ←→ (TLR) থেকে
[theke]
```

The choice of inflection depends on the spelling of the translated reference object as said before. If the translated reference object ends with a vowel, যের is added to it; else if ends with a consonant, ের (*er*)is added to it; otherwise (it ends with a *matra*) র (*r*)is appended with it. The postpositional word is placed after the inflected reference object in Bengali. The choice of the postpositional word depends on the semantic information about the reference objects as collected from the *WordNet*. In cases with one postpositional word, there is no need to know the semantic features of the reference objects. For example, '*since*', as a preposition, is always translated as থেকে (*theke*) in Bengali, irrespective of the reference object. Again in some cases, this semantic information about the reference object does not suffice to translate the preposition properly.

Consider the following examples that include the preposition *before* in two different senses.

```
(1)  He  stood  before  the
door. ←→ সে দরজার সামনে দাঁড়াল
(se [he] darja-r samne [the
door    before]    dandalo
[stood])
(2) He reached  before  eve-
ning. ←→  সে  সন্ধ্যার  আগে
পৌঁছাল (se [he] sondhya-r age
[evening before] pouchhalo
[reached])
(3) He reached  before John.
←→ সে জনের আগে পৌঁছাল (se [he]
jan-er age [John before]
pouchhalo [reached])
```

From the *WordNet*, the system acquires the semantic information that 'door' is a hyponym of

'artifact', whereas 'evening' and 'me' (which represents a person) are not. Thus 'with' is translated to Bengali as - র সামনে in sentence (1), and takes the meaning - (র / ের / যের) আগে in sentence (2) and (3).

As there is no ambiguity in the meaning of compound prepositions and phrase prepositions, a simple listing of them (along with their Bengali representations) suffices to translate them. We have prepared a list that contains the phrase prepositions and compound prepositions in English along with their Bengali translations.

English	Bengali
in spite of	সত্ত্বেও [*satteo*]
away from	- (র / ের / যের) থেকে দূরে [- (*r / er / yer*) *theke dure*]
owing to	- (র / ের / যের) কারণে [- (*r / er / yer*) *karane*]
apart from	ছাড়াও [*chhadao*]
Instead of	- (র / ের / যের) পরিবর্তে [- (*r / er / yer*) *paribarte*]
along with	- (র / ের / যের) সাথে [- (*r / er / yer*) *sathhe*]

5.3   Translation of English Idiomatic PPs

The meaning of an idiomatic PP cannot be derived from the meanings of its components. The simplest way to tackle them is to maintain a listing of them. A list or a direct Example Base is used which contains idioms, which start with prepositions, along with their Bengali translations. Such an idiom is treated like any other PP during the word-reordering phase. Here are some examples of them:

```
(1)  at times ←→  সময়ে সময়ে
(samaye samaye)
(2) by hook or crook  ←→
যেভাবেই হোক (jebhabei hok)
(3) to a fault ←→ মাত্রাতিরিক্ত
(matratirikto)
```

6   Conclusion

In the present study, the handling of English prepositions in Bengali has been studied with reference to a machine translation system from English to Bengali. English prepositions are han-

dled in Bengali using inflections and / or using post-positional words. In machine translation, sense disambiguation of preposition is necessary when the target language has different representations for the same preposition. In Bengali, the choice of the appropriate inflection depends on the spelling of the reference object. The choice of the postpositional word depends on the semantic information about the reference object obtained from the *WordNet*.

Acknowledgements

References

Alam, Yukiko Sasaki. 2004. Decision Trees for Sense Disambiguation of Prepositions: Case of Over. In *HLT/NAACL-04*.

Brala, Marija M. 2000. *Understanding and translating (spatial) prepositions: An exercise in cognitive semantics for lexicographic purposes*.

Brugman, Claudia. 1988. *The story of over: Polysemy, semantics and the structure of the lexicon*. New York: Garland Press. [1981. The story of over. Berkely, CA: UC-Berkely MA thesis.]

Emonds, Joseph. 1985. *A unified theory of syntactic categories*. Dordrecht: Foris.

Fellbaum, Christiane D. ed. 1998. *WordNet – An Electronic Lexical Database,* MIT Press, Cambridge, MA.

Fauconnier, Gilles. 1994. *Mental spaces*. Cambridge: Cambridge University Press.

Herskovits, Annette. 1986. *Language and spatial cognition An interdisciplinary study of the prepositions in English*. Cambridge: Cambridge University Press.

Hill, Cliffort 1982. *Up/down, front/back, left/right. A contrastive study of Hausa and English*. In Weissenborn and Klein, 13-42.

Jackendoff, Ray. 1977. *The architecture of the language*. Cambridge, MA: MIT Press.

Lakoff, George and Mark Johnson. 1980. *Metaphors we live by*. Chicago: University of Chicago Press.

Langacker, Ronald. 1987. *Foundations of cognitive grammar*, vol. 1. Stanford, CA: Stanford University Press.

Lindstromberg, Seth. 1997. *English prepositions explained*. Amsterdam: John Benjamins.

Naskar, Sudip Kr. and Bandyopadhyay. Sivaji. 2005. A Phrasal EBMT System for Translating English to Bangla. In *MT Summit X*.

Pullum, Geoffrey and Rodney Huddleston. 2002. *Prepositions and prepositional phrases*. In Huddleston and Pullum (eds.), 597-661.

Rauh, Gisa. 1993. *On the grammar of lexical and nonlexical prepositions in English*. In Zelinskiy-Wibbelt (eds.), 99-150.

Sopena, Joseph M., Agusti LLoberas and Joan L. Moliner. 1998. A connectionist approach to prepositional phrase attachment for real world texts. In *COLING-ACL '98*, 1233-1237.

Tezuka, Taro, Ryong Lee, Yahiko Kambayashi and Hiroki Takakura. 2001. Web-based inference rules for processing conceptual geographical relationships. *Proceedings of Web Information Systems Engineering*, 14-21.

Tyler A. and Evans V. 2003. *Reconsidering prepositional polysemy networks: the case of over**. In B. Nerlich, Z. Todd, V. Herman, & D. D. Clarke (Eds.), Polysemy: Flexible patterns of meanings in mind and language, pp. 95-160. Berlin: Mouton de Gruyter.

Visetti, Yves-Marie and Pierre Cadiot. 2002. *Instability and the theory of semantic forms Starting from the case of prepositions*. In Feigenbaum, Susanne and Dennis Kurzon (eds.), 9-39.

Voss, Clare. 2002. *Interlingua-based machine translation of spatial expressions*. University of Maryland: Ph.D. Dissertation.

Wahlen, Gloria. 1995. *Prepositions illustrated*. Michigan: The University of Michigan Press.

Xu, Yilun Dianna and Norman Badler. 2000. Algorithms for generating motion trajectories described by prepositions. *Proceedings of Computer Animation 2000*, 30-35.

Yates, Jean. 1999. *The ins and outs of prepositions A guidebook for ESL students*. New York: Barron's.

Yeh, Alexander S. and Marc B. Vilain. 1998. Some properties of preposition and subordinate conjunction attachments. In *COLING-ACL '98*, 1436-1442.