Transferring of Prepositional phrases in English into Tamil: A machine learning approach

TRANFERING OF PREPOSITIONAL PHRASES IN ENGLISH INTO TAMIL: A MACHINE LEARINING APPROACH

S. Rajendran , M. Anand Kumar, Soman K.P

Amrita Vishwa Vidyapeetham,

Coimbatore

1. Introduction

Ambiguity is the core problem in computational linguistics. Ambiguity arises at almost every level of language processing, from word level processing tasks like part of speech tagging to high-level tasks like discourse planning. A machine has to be able to resolve these ambiguities and combine the information from different levels into an unambiguous meaning representation in order to understand a human language.

Preposition is one of the word classes which is both frequent and highly ambiguous. The different senses of a preposition express different relations between the preposition complement and the rest of the sentence. The sense is related to the semantic role of the dominating prepositional phrase. To understand the semantics of a prepositional phrase, a system would need access to both the higher level semantics of the semantic role and the finer word-token level semantics of the preposition.

Prepositions are not given the attention they deserve in the earlier studies on resolution of ambiguity. Even in lexicographic works including dictionaries, prepositions are not elaborately discussed explicating the ambiguity they carry along with them. Presuppositions are not deeply studied in the corpus analysis unlike the other parts of speech. Though prepositions are only a closed set of words exhibiting certain grammatical functions, their polysemous nature is comparable to other parts of speech. Similar to the major parts of speech like noun and verb, prepositions also creates problem in their interpretation. The interpretation of prepositions becomes a challenge to the computational community who are involved in natural language processing. They are closely related to verbs as the indicators of their internal arguments.

Preposition is a term used in grammatical classification of words referring to the set of items which typically precede NP (often single nouns or pronouns) to form a single constituent of structure. Prepositions normally precede nouns or pronouns. For example

The cow is grazing in the field.

The preposition 'in' shows the relationship between *cow* and *field*. In the above sentence object of the preposition *field* comes after the preposition *in*. Hence the noun or pronoun which is used with a

preposition is called its object. It is in the accusative case and is said to be governed by the preposition. In the above given sentence the noun *field* is in accusative case and is governed by the preposition in. A preposition may have two or more objects as in the following sentence.

The road runs over hill and plain.

It has to be noted here that prepositions can also be an adverb and that is they can be used without an object. If personal pronouns *I, we, he, she, they* etc are used as the object of a preposition, then their objective form *me, us, him, her, them*, have to be used.

Tamil makes use of postpositions instead of prepositions; these postpositions could be affixes or free forms or a combination of both.

avan meecai meel puttakattai vaittaan

he table on book-ACC keep-PAST-3PS

'He kept the book on the table'

Though we can draw one to one correspondence between English prepositions and Tamil postpositions in many instances, there are drastic differences between them in some instances.

This paper aims at transferring of prepositional phrases in English into Tamil by making use of a machine learning approach. We are considering only two prepositions 'for' and 'with' to substantiate our arguments.

2. Ambiguity in prepositions

Actually prepositions themselves are ambiguous at the source language level. One can expect 'transfer ambiguity' at the transfer level when we move from English to Tamil. Take for example the following sentences with the preposition 'for'.

1. He boarded the train for Jaipur.
2. I waited for you.
3. Ram has sympathy for the poor.

Look at their translations into Tamil.

1a. *avan jaipUrukku rayil ERinAn*

2a. *wAn unakku vENti/Aka kaththiruwthEn*

3b. *rAmukku EzaikaL itam irakkam irukkiRathu.*

In the first sentence the preposition 'for' is matched to dative suffix –*ukku* in Tamil; in the second sentence 'for' is matched against the benefactive *vENti/Aka* and in the third sentence 'for' is matched with the receiver postposition *iTam* 'to'. This three way distinction of 'for' in Tamil exhibit the three-way ambiguity at the source language which is reflected at the transfer level.

Now let us look at the sentences with 'with' preposition.

4. He apologized with her.
    4a. *avan avaL-itam mannuppu kEttAn*

5. Blend water with milk.
    5a. *taNNIrutan pAlaik kala*
6. We walk with legs.

    6a. *wAm kAlkaLAl natakkinROm*

In the first sentence the preposition 'with' is matched with the addressee marker *itam* in Tamil; in the second sentence 'with' is matched against the sociative postpostion *utan* and in the third sentence 'with' is matched with the instrumental case suffix *Al* 'by'. This three way distinction of 'with' in Tamil exhibit the three-way ambiguity at the source language which is reflected at the transfer level.

Proposition sense disambiguation requires lexical and world knowledge, and contextual information as generally with word sense disambiguation. For example, consider the sentences "Ram built a house *with* a magic wand" and "Ram built a house *with* a bed room". In the first example, '*with*' is used to indicate the *instrument* that is used in building, while in the second example '*with*' is used to indicate the part of a house that has a bed room built-in. To distinguish these two senses of '*with*', we need to understand the difference between 'magic wand' and 'bed room'. This requires world knowledge or the understanding of word meanings.

3. *Earlier works*

A number of researchers including Alam (2004), Harabagiu (1996), O'Hara and Wiebe (2002, 2003), Sopena, LLoberas, and Moliner (1998), Sablayrolles (1995), Saint-Dizier and Vazquez (2001), Litkowski (2002), and Boonthum, Toida and Levinstein have studied recently disambiguation of the preposition. Alam (2004) studied the disambiguation of the preposition 'over' by taking into account two things: one is the meaning expressed by the complement noun phrases of the prepositions and the other is the head of the prepositional phrases (verb or noun phrases). Harabagiu (1996) made use of WordNet to

disambiguate prepositional phrase attachment. A special issue of *Computational linguistics* (Baldwin et al., 2009) was devoted to discuss about the issues on preposition. Preposition sense disambiguation was one of the SemEval 2007 tasks (Litkowski and Hargraves, 2007), and was subsequently explored in a number of papers using supervised approaches. O'Hara and Wiebe (2009) presented a supervised preposition sense disambiguation approach which explores different settings. Tratz and Hovy (2009), and Hovy et al. (2010) make explicit use of the arguments for preposition sense disambiguation, using various features. Rudzicz and Mokhov (2003) and O'Hara and Wiebe (2003) have studied the constraints of prepositional constructions to annotate the semantic role of complete prepositional phrases with FrameNet and Penn Treebank categories.

The present study is much rarer of the kind of studies mentioned above as it aims to resolve the prepositional ambiguity at the transfer level.

4. *Methodology*

English sentences with the prepositions "*for*" and "*with*" are collected and manually tagged with equivalent postpositions in Tamil. We devided the annotated English preposition corpus into train and test sets. The training dataset consists of 430 sentences and testing set has 100 sentences Sample English sentences with preposition tag are given in Table.1. The preposition tag contains the Tamil postposition information. By using this method, the prepositions of English are disambiguated and translated into Tamil postpositions.

To disambiguate a preposition p, our system uses the bag-of-words and linguistic information as features. The essential linguistic information for each English sentence is extracted using the Stanford parser toolkit. This linguistic information such as lemma, POS tag, and dependency tag is used as features in machine learning based disambiguation model. For training the model we utilized SVM-Light, a public distribution of SVM (Support Vector Machines) by (Joachims, 1999). The preposition disambiguation framework is shown in the Figure.1.

Table:1 Sample English Sentences and its Prepositional Tag

S.No	English Sentences	Preposition Tag
1	Dussehra is celebrated _for_ ten days.	NUL
2	I have played outside _for_ an hour.	DAT-Aka
3	She does not bring water _for_ me.	DAT
4	I had headache _for_ two or three days.	Aka
5	His body is covered _with_ hair.	Al
6	Peter has fallen out _with_ his boss.	il-iruwthu
7	I have lived _with_ my parents for over 10 years.	Aka
8	She did not come _with_ me.	utan
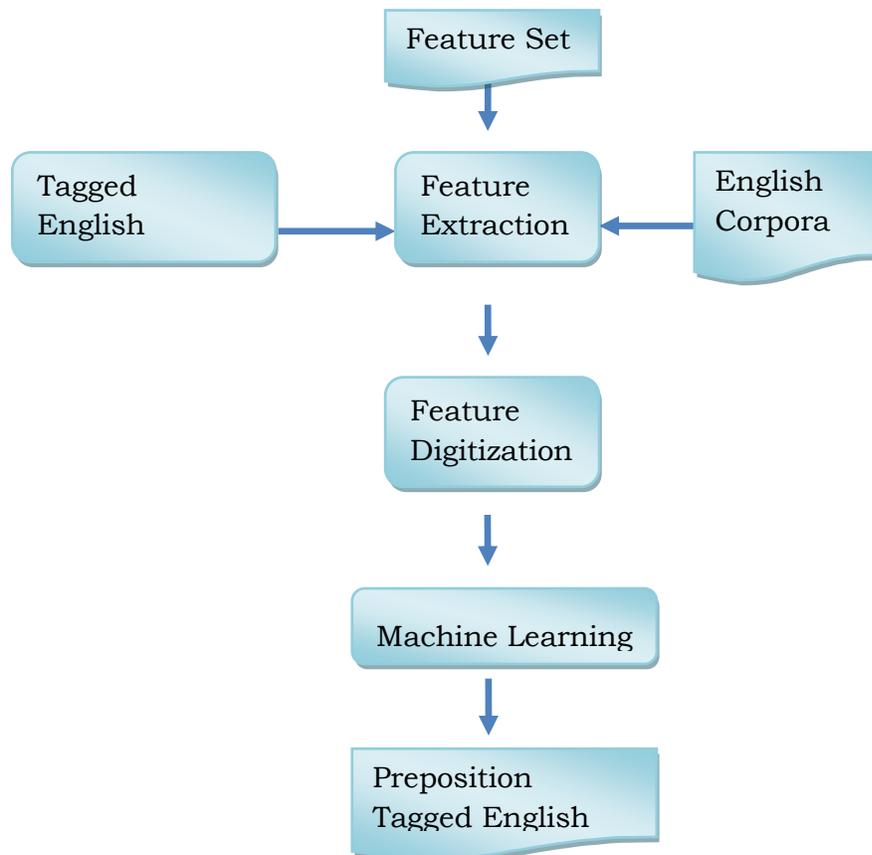
Figure: 1 Framework for English Preposition disambiguation

Our investigation revealed that machine learning method for solving the preposition disambiguation problem using existing knowledge sources as features improved the performance of preposition translation. In future, the feature set can be further tuned with the use of English Word Net and the Semantic role's of verb in the sentences.

5. *Conclusions*

 It can be inferred by linguistic analysis that within the context for prepositions, the governor (head of the NP or VP governing the preposition), the object of the preposition (i.e., head of the NP to the right), and the word directly to the left of the preposition have the highest influence. While disambiguating prepositions the maximal accuracy can be achieved by considering the context, features, and granularity. Though the preliminary result is encouraging, various issues still need to be addressed, *i.e.* improving the result by disambiguating noun senses or using world knowledge or context information.

Preposition sense disambiguation has many potential uses. For example, due to the relational nature of prepositions, disambiguating their senses can help with all-word sense disambiguation. In machine translation, different senses of the same English preposition often correspond to different translations in the foreign language. Thus, disambiguating prepositions correctly may help improve translation quality.

6. *References*

1.   Alam, Y. 2004. Decision trees for Sense Disambiguation of Preposition: Case over. In HLT-NAACL, Computational Lexical Semantics Workshop, Boston: MA, pp 52-59.

2.   Baldwin, T. V. Kordoni, and A. Villavicencio. 2009. Prepositions in applications: A survey and introduction to the special issue. Computational Linguistics, 35(2):119–149.

3.   Bannard, C and T. Baldwin. 2003. Distributional Models of Preposition Semantics. In ACL-SIGSEM, Workshop on the Linguistic Dimensions of Prepositions and their use in Computational Linguistics Formalism and Applications, Toulouse: France, pp 169-80.

4.   Boonthum C. S. Todia and I. Levistein. 2005. Sense Disambiguation of preposition 'with'. Department of Computer Science, Old Dominion University, USA.

5.   Dorr, Bonnie 1992. The Use of Lexical Semantics in Intelingual Machine Translation,   Machine Translation, 7:3, pp 135-193.

6.   Harabagiu S. 1996. An Application of WordNet to Prepositional Attachment. In ACL, Santa Cruz, pp. 360-363.

7.   Hovy, D. S. Tratz, and E. Hovy. 2010. What's in a Preposition? Dimensions of Sense Disambiguation for an Interesting Word Class. In Coling 2010: Posters, pages 454–462, Beijing, China, August. Coling 2010 Organizing Committee.

8.   Joachims, T. 1999. Transductive Inference for Text Classification using Support Vector Machines. International Conference on Machine Learning (ICML), 1999.

9.   Levin, B. 1993. English Verb Classes and Alternations: A Preliminary Investigation, University of Chicago press, Chicago; IL.

10. Litkowski, K. 2002. Digraph Analysis of Dictionary Preposition Definition. In ACL-SIGLEX, SENSEVAL Workshop on Word Sense Disambiguation: Recent Success and Future Directions, Philadelphia: PA, pp. 9-16.

11. Litkowski, K. and O. Hargraves. 2007. SemEval-2007 Task 06: Word-Sense Disambiguation of Prepositions. In Proceedings of the 4th International Workshop on Semantic Evaluations (SemEval-2007), Prague, Czech Republic.

12. O'Hara T. and J. Wiebe. 2002. Classifying Preposition Semantic Roles using Class-based Lexical Associations. Technical Report NMSU-CS-2002- 013, Computer Science Department, New Mexico State University.

13. O'Hara, T. and J. Wiebe. 2003. Preposition semantic classification via Penn Treebank and FrameNet. In Proceedings of CoNLL, pages 79–86. O'Hara, T and J. Wiebe. 2009. Exploiting semantic role resources for preposition disambiguation. Computational Linguistics, 35(2):151–184.

14. Rudzicz, F and S. A. Mokhov. 2003. Towards a heuristic categorization of prepositional phrases in English with wordnet. Technical report, Cornell University, arxiv1.library.cornell.edu/abs/1002.1095-?context=cs.

15. Saint-Dizier, P and G. Vazquez. 2001. A Compositional Framework for Prepositions. In ACLSIGSEM, International Workshop on Computational Semantic, Tilburg: Netherlands.

16. Sopena, J. A. LLoberas, and J. Moliner. 1998. A Connectionist Approach to Prepositional Phrase Attachment for Real World Text. In ACL, Montreal, Quebec: Canada, pp. 1233-1237.

17. Sablayrolles, P. 1995. The Semantics of Motion. In EACL, Toulouse: France, pp. 281-283.

18. Tratz. S. and D. Hovy. 2009. Disambiguation of Preposition Sense Using Linguistically Motivated Features. In Proceedings of Human Language Technologies: The 2009 Annual Conference of the North American Chapter of the Association for Computational Linguistics, Companion Volume: Student Research Workshop and Doctoral Consortium, pages 96–100, Boulder, Colorado, June. Association for Computational Linguistics.

✿✿✿