**English to Chinese translation of prepositions**
Li, Hui; Japkowicz, Nathalie; Barrière, Caroline

National Research Council Canada    Conseil national de recherches Canada

Canada

National Research
Council Canada

Conseil national
de recherches Canada

Institute for
Information Technology

Institut de technologie
de l'information

# NRC·CNRC

*English to Chinese Translation of
Prepositions ***

Li, H., Japkowicz, N., Barrière, C.
May 2005

Canada

# English to Chinese Translation of Prepositions

Hui Li, Nathalie Japkowicz, Caroline Barrière

School of Information Technology and Engineering, University of Ottawa, Canada

Abstract. Machine translation of prepositions is a difficult task; little work has been done, to date, in this area. This article suggests addressing the problem using a semantic framework for the interpretation of the surrounding elements of a preposition in the source language. This framework, called Use Types, will reduce the set of possible prepositions in the target language, therefore helping the translation process. This approach is not language dependent, but we focus, here, on English and Chinese, and we also specifically look at three prepositions: in, on and at. The article describes machine learning experiments designed and conducted in which WordNet is employed to lead to an automatic discovery of the Use Types. Results are analyzed and discussed and a practical use of the system is suggested along with the preliminary results it obtains.

## 1    Translation of Prepositions: Looking into Use Types

Prepositions play a very important role in language. Without or with wrong prepositions, sentences are difficult to understand. Translation of prepositions is difficult and little research has been done on it compared to work done on other words. Furthermore, the issue received close to no attention in the context of English to Chinese. Although some automated translation systems, e.g. Worldlingo[1] are developed, prepositions are sometimes translated in a non-colloquial or non-understandable fashion. Among the 300 examples we collected from "The Bible" [8], 103 of them were meaningfully translated by Worldlingo, but 197 examples were translated in a non-understandable manner. The problem of translation of prepositions is twofold. First, high usage of prepositions unfortunately comes with a high degree of polysemy; and meanings in different languages do not necessarily match. Second, even for a single meaning, different prepositions are possible.

Our hypothesis turns toward work on conceptualization [2] which suggests an interpretation of a preposition based on the semantic interpretation of the nouns surrounding it. This

---

[1] Worldlingo can be found at www.worldlingo.com

hypothesis is grounded in earlier work by Japkowicz [4,5] in which differences between English and French locative prepositions were analyzed based on the observation that these two languages sometimes conceptualize objects in a different way. As an example, consider the object *bus*. A *bus* has a roof and several sides, so it can be conceptualized as container as in French. However, a *bus* also has a platform which can be seen as playing a more important role than the roof or the sides, resulting in a conceptualization as a surface as in English. In our work, a similar idea of conceptualization of objects is explored but through the use of Use Types, as developed by Herskovits [3], which correspond to patterns of a set of sentences from the perspective of cognitive science. Herskovits summarized a list of Use Types for each preposition. In the present work, Use Types are adapted and their range extended outside of locations to include other situations, like time, state, and direction. Table 1 shows a sample of the Use Types for preposition in. The Use Types developed in this research are indicated by a "*". For the most part, we can see that a single Use Type corresponds to a single Chinese meaning, although in some cases, a few different Use Types may belong to the same Chinese meaning, as in the first four examples.

**Table 1**. a sample of the Use Types for preposition in.

| Use Types | Example of Sentence | Chinese meaning |
|---|---|---|
| Spatial entity in container | The preserves in the sealed jar | （指地点）在…中；在…内;在…上 |
| Physical object "in the air" | The bird in the air | |
| Physical object in the roadway | The ruts in the road | |
| Person in institution | A man in a red hat | |
| Person in clothing | A man in a red hat | （指衣服等）穿着，戴着 |
| * Physical object in situation, or state | They fell in love. | （表示情况或状态）在…状态中 |
| * Physical object in environment | She is standing outside in the cold. | （表示环境或境遇等）在…环境下 |
| * person in career, activity | He's in the army | （表示职业，活动） |
| * object in direction | He could number the fields in every direction | （指方向）在…方向 |
| * Physical object in the time span it takes to finish the described action | I will be back in a short time | （指时间）过（若干时间），在（若干时间）内 |
| * Physical object in shape, form, order | words in alphabetical order | （表示形式，形状，排列） |
| * Object in way, medium, tool, or material. | A message in code. | （表示表达的方法，媒介，工具，原料等） |

The challenge then remains to automatically extract the Use Type from an English sentence, from which we will obtain the corresponding Chinese meaning (as found in a dictionary), thus leading to a reduced set of possible Chinese prepositions.

In the present work, we intend to use ML techniques to discover the Use Type directly from a preposition in context. Inspired by [5], we rendered some Use Types more specific, aiming at

finding middle ground of generality/specificity that would make the Use Types useful as semantic interpretations for translation in different languages.

## 2     Experimentation & Results

To categorize an English sentence into a Use Type, the first step is to generalize the nouns surrounding the preposition to conceptual levels appropriate in Use Type definitions. For example, the noun *farmer,* present in a sentence should be generalized to its superclass person which could be part of a Use Type. Such relations can be found in a lexical knowledge base, such as WordNet [1,7], containing information organized as a lexical hierarchy. In more details, all of the nouns in WordNet are organized into synsets, which, in turn, are organized into hierarchies. We therefore design an experiment to make use of WordNet in the semi-automatic determination of the Use Type that will correspond to a preposition in context.

In order to translate prepositions from English to Chinese, we followed the following steps:
- Gather a corpus of English sentences with their Chinese translation
- Shallow-parse the sentences to extract the nouns around the prepositions: reference and located object
- Find the nouns' hypernyms using WordNet
- Use these hypernyms, together with the preposition, as features for ML training set.
- Label each training example with its class.
- Train some classifiers on the data gathered in the previous phases.

We used two labeling strategies to test whether Use Types are needed or not:
- Experiment 1 : Use the 62 Use Types as the classes to be learned
- Experiment 2 : Use the 74 Chinese prepositions directly without Use Types.

Table 2 show the result of learning by Use Type vs. by Chinese preposition. It displays error rates of classification and only shows the results of those classifiers that performed relatively better, where C4.5 is a decision tree learner, and PARTruleLearner is a learner that build rules from partial decision trees. Sentences used in the experimentation come from dictionaries [9, 10], Herskovits's book [3], the online corpus *The Little Prince* [11], Jane Austen's *Pride and Prejudice*, and HongKong Polytechnic University's online Magazine Articles. The total number of instances used is 2000, and we conducted a 10-fold Cross-validation testing policy. We also calculated and showed Baselines at the bottom of the table based on the following three ways. First, we randomly select a Use Type or a Chinese preposition (Baselines 1). Second, we choose the most frequent Use Type or Chinese preposition all the time (Baselines 2). Third, we randomly select a Use Type or Chinese preposition according to the probability

that each Use Type or preposition is chosen (Baselines 3). Several other experiments are also reported in [6].

The best result we obtained is 30.6733% in the case of preposition *at* when learning by Use Type. The worst result is 66.2368% in the case of preposition *in* when learning by Chinese preposition directly. The baseline of the most frequent Use Type is as high as 96.95%, and the baseline of the most frequent Chinese preposition is 84.85%. Furthermore, we found that, in general, the difference in performance of Usetypes versus Chinese prepositions directly can go as high as 20%. These results demonstrate the utility of using Use Types.

**Table 2.** Error rate of learning with each preposition separately

| Classifier | At | | In | | On | |
|---|---|---|---|---|---|---|
| | Use Type | Cprep. | Use Type | Cprep. | Use Type | Cprep. |
| C4.5 | 30.6733 | 48.8778 | 42.0819 | 58.2206 | 55.814 | 62.4313 |
| PARTruleLearner | 30.9227 | 48.3791 | 41.2811 | 61.8683 | 56.4482 | 66.2368 |
| Baseline 1 | 99.45 | 99.8 | 95.5 | 99.7 | 97.35 | 99.75 |
| Baseline 2 | 95.05 | 84.85 | 90.15 | 95.25 | 96.5 | 92.4 |
| Baseline 3 | 99.6 | 99.75 | 96.15 | 99.65 | 97.15 | 99.8 |

## 3    Practical Application of this Research

The practical purpose of our research was to build a post-processing unit that would correct the preposition output by the automated translator when necessary. To assess the accuracy of that unit, we use 300 examples from "the Bible", which were not used in previous experiments, and use these data only as test sets. We ran Worldlingo and our technique on these examples separately.

Our results are as follows: Among the 103 examples that Worldlingo translated understandably, 39 of them were wrongly translated by our system. However, among the 197 examples that were wrongly translated by Worldlingo, 108 of them were translated meaningfully by our system. This means that our post-processing unit allowed us to improve the output of Worldlingo on prepositions "in", "on" and "at" by 23%, bringing it from an accuracy of 34.33% to an accuracy of 57.33%. For example, the sentence *the man in red* was translated to 人在红色 in Chinese, which means *man at red* in English. While our approach will get the Use Type of *person in clothing,* which corresponds to the Chinese Meaning of（指衣服等）穿着，戴着, which means *wear* in English. The result suggests that our approach is a valuable addition to existing well-recognized translation system.

# 4    Conclusion & Future Work

The purpose of this paper was first to present Use Types as a possible semantic interpretation framework for prepositions in context, with a purpose of machine translation. We referred to previous related work by Herskovits, and Japkowicz, who focused on locative prepositions and broadened their work by expanding to Use Types for non-locative prepositions.

The experiments we conducted showed that introducing Use Types as an intermediate step can help improve the accuracy of translation.  Furthermore, we found that Wordnet along with Machine Learning tools could be useful in the automatic assignation of a Use Type for a preposition. Our approach is also valuable practically as we showed that combined our approach to the output of WorldLingo lead to a non-negligible accuracy improvement of 23%.

For future work, we need to collect more data to better evaluate this approach.  We also expect to apply this approach to other prepositions, and languages. Lexical resources other than Wordnet also need to be investigated.

## References
1. Christiane Fellbaum, editor. 1998 *WordNet: An Electronic Lexical Database*. The MIT Press.  23-46
2. M. Grimaud. 1988 Toponyms, Prepositions, and Cognitive Maps in English and French. *Journal of the American Society of Geolinguistics*, vol. 14, pp. 5476.
3. A. Herskovits. 1986. *Language and spatial cognition: an interdisciplinary study of the prepositions in English*. Cambridge [Cambridgeshire]; New York: Cambridge University Press. 39-54, 86-94, 127-155
4. N. Japkowicz, J. M. Wiebe.  1991. A System for Translating Locative Prepositions from English into French.  *29th Annual Meeting of the Association for Computational Linguistics* 29 (18-21 June):153-160.
5. Nathalie Japkowicz. 1990. *The Translation of Basic Topological Prepositions from English into French*. M.S. Thesis, published as Technical Report CSRI243, University of Toronto.
6. Hui Li, 2004. *Use Types for English to Chinese Translation of Prepositions*. M.S. Thesis, published as Technical Report, University of Ottawa.
7. Miller. 1990. WordNet: An On-line Lexical Database. In *International Journal of Lexicography*, 3(4)
8. *The Bible*. http://www.bible.org/page.asp?page_id=2485
9. *Oxford Advanced Learner's Dictionary of Current English With Chinese Translation*. The Commercial Press. Oxford University Press. 1994.  65-66, 573-574, 782-783
10. *The Advanced Learner's Dictionary of Current English With Chinese Translation*. Oxford University Press. 1978. 58, 536-538, 734-735
11. *The Little Prince*. The Hong Kong Polytechnic University. http://www.engl.polyu.edu.hk/tricorpus/