Target Language Preposition Selection – an Experiment with Transformation-Based Learning and Aligned Bilingual Data

Ebba Gustavii

Department of Linguistics and Philology,
Uppsala University, Sweden
ebbag@stp.ling.uu.se

Abstract. The translation of prepositions is often considered one of the more difficult tasks within the field of machine translation. We describe an experiment using transformation-based learning to induce rules to select the appropriate target language preposition from aligned bilingual data. Results show an accuracy of 84.9%, to be compared with a baseline of 75.5%, where the most frequent translation alternative is always chosen.

1. Introduction

The selection of prepositions may be due to lots of factors, some of which are mainly idiosyncratic to the language in question, and some of which are dependent on the content that the prepositions contribute with. In the field of machine translation, the translation of prepositions is thus often considered to be one of the more difficult issues, and often there are separate modules dedicated to that task.

The many dependencies, often lexical in nature, make it cumbersome, maybe even unfeasible, to manually identify and formalize the constraints necessary to translate prepositions appropriately. With the growing bulk of large parallel corpora, however, supervised machine-learning techniques may be used to facilitate the tedious work: either by revealing patterns hidden in the data, or more directly, by using the techniques to generate classifiers selecting the appropriate preposition.

Here we will take the latter approach, and apply transformation-based learning to induce rules for correcting prepositions output by a rule-based machine translation system. Selectional constraints will be sought in the target language context. For training, however, solely aligned bilingual corpus data will be used, and one rule sequence will be induced for each source language preposition. Each classifier will be trained on target language prepositions actually being aligned to the respective source language preposition.

The paper is organized as follows: In the second section, we will look into the heterogeneous nature of prepositions and discuss some of its implications on the translation process. In the third section, we will briefly review some previous experiments on related tasks; we will specifically consider whether they have involved the use of aligned bilingual data or not. The fourth section will outline and motivate the main features of the current approach. In the fifth section, transformation-based learning will be introduced. The sixth section presents the actual experiment: the data and tools, the parameter settings and the choice of templates. Section seven is devoted to a presentation of the results. In the final section, some concluding remarks will be given.

2. How Prepositions Translate

Linguists often distinguish two types of prepositional uses; their functional use and their lexical use.[1] In its functional use, a preposition is governed by some other word, most often by a verb as in example 1, but sometimes by an adjective (*afraid of*), or a noun (*belief in*).

1. I believe in magic.

---

[1] Other labels that have been used for approximately the same distinction are: determined vs. non-determined, synsemantic vs. autosemantic and non-predicative vs. predicative. (Tseng, 2002)

The selection of a functional preposition is determined by the governor, and the preposition is typically not carrying much semantic information. This is evident when comparing semantically similar verbs taking different prepositions, such as *charge* NP *with* NP, *blame* NP *for* NP, and *accuse* NP of NP. When translating a functional preposition, the identity of the source language preposition is thereby of less importance. Rather, the crucial information lies in the co-occurrence patterns of the target language.[2] Working from an interlingual perspective, Miller (1998) suggests that content-free prepositions, which roughly coincide with prepositions in their functional use, need not be represented at the inter-lingual level at all, but are better treated as a problem of generation. Within a corpus-based strategy, this would correspond to using only monolingual target data as corpus data.

In their lexical use, prepositions are not determined by some governing word, but are selected due to their meaning. In example 2, other prepositions than *in* are grammatically valid, e.g. *under* or *beside*, but these would alter the meaning of the utterance.

2. The rabbit is in the hat.

When translating a lexical preposition, the identity of the source language preposition, or rather the content it carries, is thus of importance; something which implies the need for bilingual data.

The best place to look for clues for the selection of a target preposition is evidently dependent on whether the source preposition is functional or lexical. The optimal strategy would thus be to treat functional and lexical prepositions differently. In practice, however, it turns out to be very difficult to classify prepositional uses into these categories. The verb *put*, for instance, subcategorizes for a direct object and a locative where the latter often is expressed by a prepositional phrase (e.g. *put the vase on the table*). The prepositional phrase is thus subcategorized for, but still, the selection of the preposition is semantically based. Moreover, lexical prepositions are not always chosen on the basis of their

content only, but may be further constrained by the nouns they govern. We say *at the bank* and *in the store*, though the prepositions contribute with approximately the same meaning in both cases. (For an in-depth discussion on classificational issues of prepositions, see Tseng (2000)).

When choosing a strategy for selecting the appropriate target preposition, one should thus keep both kinds of prepositional uses in mind - something which implies the need for both bilingual and monolingual data.

3. Related Work

Several strategies have been suggested for the task of selecting the appropriate target word in context. Most of these, however, address the translation of content words. We will take a brief look at some of the more influential such proposals. For the specific task of selecting the appropriate target preposition, we will take a closer look at a strategy proposed by Kanayama (2002).

The methods suggested for target word selection may be classified according to whether they make use of aligned bilingual corpus data or not.

The obvious advantage of not using aligned bilingual corpora, but monolingual corpora instead, is the vast increase in data available. Dagan and Itai (1994) suggest a statistically-based approach using a monolingual target corpus and a bilingual dictionary. When the bilingual dictionary gives several translation alternatives for a word, the context is considered, and the alternatives are ranked according to how frequently they occur in a similar context in the target language corpus. When there is more than one selection to be made, the order is determined by a constraint propagation algorithm. The results taken from an evaluation on a small English-Hebrew test set were promising, showing a recall of 68% and a precision of 91%.

Kanayama (2002) presents an algorithm specifically tailored to acquire statistical data for the translation of the Japanese postposition *de* to the appropriate English preposition. Following Dagan and Itai (1994), he selects the target word on the basis of co-occurrence patterns in the target language. For the experiment, however, also a Japanese parsed corpus is used, from which almost half a million verb phrases with the postposition *de* are extracted. These are par-

---

[2] This is a bit simplified. The particular syntactic relation that is signaled by the source language preposition may of course be of relevance.

tially translated to English, with the preposition left unspecified. Next, a parsed English newspaper corpus is searched for the partial translations where the unspecified preposition is instantiated as one of six predefined translations of *de*. When translating *de*, the most frequent target preposition, given the surrounding verb and noun, is chosen. In case there are no such tuples in the data, only the noun context is considered. As a last resort a default preposition is selected. The reported total precision was 68.5%, to be compared with a baseline of 41.8% (where the default translation is always chosen).

Dagan and Itai (1994) note that the use of non-aligned corpus data alone, makes it impossible to distinguish between instances of a target word that corresponds to different source words when gathering context statistics for the target words. Therefore, each instance of a target word will be treated as a translation of all the source words for which it is a potential translation. In both experiments, this has been reported to be a source of errors. For instance, the algorithm suggested by Kanayama selects *with* over *for* in *work* (*with/for*) *the company*, since that construction is the most frequent one in the target language corpus. In the particular context though, *with* is not an appropriate translation of *de*, but corresponds to the translation of some other adposition.

Approaches to target word selection that make use of aligned bilingual data have also been suggested. Among the more influential ones are Brown et al (1991a; 1991b). In their proposal, the translation process is preceded by a sense-labeling phase, where ambiguous words are labeled with senses that correspond to different translations in the particular target language. A word token is sense-labeled by reference to a single feature in its context (e.g. the first verb to its right). For each ambiguous word the algorithm identifies the informant site that partitions the tokens in a way that maximizes the mutual information between the senses and the aligned translations. For instance, when translating the French verb *prendre* to English, the most informative feature was found to be the accusative object (approximated as the closest succeeding noun). By incorporating the sense-labeling technique into a statistical machine translation system, Brown et al (1991b) increased

the number of acceptable produced by the system from 37 to 45 sentences out of 100. (Brown et al, 1991b)

In statistical machine translation, aligned bilingual data plays a major role in the selection of target words. Probability estimates are extracted from a translation model and a language model, which are built from an aligned bilingual corpus and a monolingual corpus, respectively. In part, however, the problem noted by Dagan and Itai (1994) still prevails; since the target language model is built on non-aligned data, there are no means to distinguish the different sources when context statistics are gathered for a target word.

4. Main Features of the Current Approach

The aim of the current experiment is to construct classifiers able to correct prepositions output from a rule-based MT-system. We will assume that the rule-based system, as a default, picks the most frequent target language preposition given the source preposition. Our task will thus be to identify the contexts where this default selection should be overridden, and the selected preposition be changed for a more appropriate one.[3] We will avoid inducing rules where a preposition should be changed to some other part-of-speech, or where it should be completely removed, since such rules would alter the output structure in an uncontrolled way. The focus will consequently be on situations where prepositions translate as prepositions. This limits the applicability of the strategy to relatively similar languages, as the ones of the current study (Swedish and English).

To induce the classifiers we will use the symbolic induction algorithm transformation-based learning (TBL) (for a very brief introduction, see section 5). TBL has successfully been applied to a wide range of NLP-tasks, e.g. part-of-speech tagging (Brill, 1995), prepositional phrase attachment (Brill & Resnik, 1994), spelling correction (Mangu & Brill, 1997) and word sense disambiguation (Lager & Zinovjeva, 2001).

---

[3] We will assume that the rule-based system annotates whether prepositions are output as defaults or have been selected by some rule. The post-processing filter should only be applied to the former ones.

For the current task, where we look for contexts in which a default selection should be overridden, we find TBL to be particularly well-suited; starting with a good heuristic and then, iteratively, define contexts where previous decisions should be changed, is at the heart of TBL.

Paliouras et al (2000) compare the performance of different machine learning techniques (symbolic induction algorithms, probabilistic classifiers and memory-based classifiers) on word sense disambiguation (WSD), and find the symbolic induction algorithms to give the best results. Since WSD and target word selection are relatively similar tasks, this gives further motivation for the choice of a symbolic induction algorithm for the task at hand.

Since the selection of target language prepositions to a great extent is due to factors idiosyncratic to the target language, we will follow Dagan and Itai (1994), and Kanayama (2002), in looking for selectional constraints in the target language context. To avoid confusing the sources, as may happen when non-aligned data is used, we will however use an aligned bilingual corpus, and induce one rule sequence for each source language preposition. Each classifier will be trained on actual translations (i.e. alignments) only of the respective source language preposition. This strategy, to look for selectional constraints in the target language context, while still keeping track of the identity of the source language preposition, may be viewed as a compromise to accommodate for both functional and lexical uses of prepositions.

The classifiers will have access to the word form, the lemma and the part-of-speech of the potential contextual triggers. We will primarily accommodate for selectional constraints triggered by governing words, or from governed nominals inside the prepositional phrase. The potential governors will be approximated as the closest preceding verb, noun or adjective, and the governed nominals, as the closest succeeding noun. With fully parsed data, the governor, as well as the governed nouns, would be recognized with higher precision. The resulting classifiers would however be dependent on having access to fully parsed data, something which is not always output from rule-based MT-systems.

5. Transformation-Based Learning

Transformation-based learning, introduced by Brill (1995), is an error-driven symbolic induction algorithm that learns an ordered set of rules from annotated training data. The format of the induced rules is determined by a set of rule templates that define what features the rules are to condition. In a first stage, the algorithm labels every instance with its most likely tag (initial annotation). It then iteratively examines every possible rule-instantiation and selects the one which improves the overall tagging the most. The iteration continues until no rule-instantiation reaches a reduction in error above a certain threshold.

In our experiments we use μ-TBL, a flexible and efficient prolog-implementation of a generalized form of transformation-based learning, developed by Lager (1999).

6. Experimental Setup

6.1. Data and Evaluation

As parallel corpus data, we have used a subset of the Swedish-English EUROPARL corpus (Koehn, n.d.). The subset consists of approximately 3

Source Language Preposition	Accuracy TBL	Accuracy Baseline	Nr of Training Instances
i (*in*)	87.0%	83.3%	27190
av (*of*)	89.4%	79.8%	21182
för (*for*)	80.2%	73.2%	14632
med (*with*)	88.6%	85.4%	8465
på (*on*)	81.1%	45.3%	7898
om (*on*)	73.4%	59.3%	7502
Total:	84.9%	75.5%	-

Table 1. Accuracy for the six most frequent source language prepositions (score threshold 2, accuracy threshold 0.6). Baseline calculated from always selecting the most frequent translation (given in brackets).

million tokens in each language, out of which approximately 90% were used for training, and the remaining 10% were left for testing. The corpus was word- aligned with the GIZA++ tool-kit (Och & Ney, 2000).

To identify the prepositions, and to accommodate for more general rules to be learnt, the corpus was part-of-speech tagged. For both languages the TnT-tagger (Brants, 2000) was used, with a model extracted from the Penn Treebank Wall Street Journal Corpus (Marcus et al, 1994) for the English part, and from the Stockholm-Umeå Corpus (Ejerhed et al, 1992) for the Swedish part (Megyesi, 2002).

In the English part, all verbs, nouns and adjectives were lemmatized with the morphological tool *morpha*. (Minnen et al, 2001)

From the aligned and processed corpus, training and testing sets were extracted for the six most frequent prepositions in the training corpus: *i*, *av*, *för*, *med*, *på* and *om*. For each of those, we extracted the aligned target language prepositions in their sentence context.

The target prepositions in the training and the testing sets were initially annotated with the most frequent translation of their respective source prepositions (as estimated from the training corpus). In so doing, we are simulating the output of an MT-system that always selects the most frequent translation of a source language preposition.

Each rule sequence was evaluated by running the built-in evaluation function in µ-TBL on its respective test set.

6.2. Templates

The templates determine the format of the rules to be learnt, or more specifically, what features should be conditioned by the rules. As was previously noted, we have defined the templates to accommodate for selectional constraints triggered either from some governing word, or from a word inside the prepositional phrase. Templates for external triggers are defined to condition the closest preceding noun, verb or adjective. There are also supplementary templates conditioning any immediately preceding word and/or part-of-speech. Templates for internal triggers are defined to condition the closest succeeding noun. Also here supplementary templates are defined to condition any immediately succeeding word and/or part-of-speech.

6.3. µ-TBL – Parameter Settings

When running the µ-TBL system, the user must decide on a minimum score threshold[4] and a minimum accuracy threshold[5]. The optimal values of these depend on the data at hand, and are best estimated empirically. Here we have only experimented with three values for each: 2, 4, and 6 as possible score thresholds, and 0.6, 0.8 and 1.0 as possible accuracy thresholds.

7. Experimental Results

The best overall results, presented in Table1, were achieved with a score threshold of 2, and an accuracy threshold of 0.6. The increase in accuracy, as compared to a baseline where the most frequent translation of each preposition is always selected, is quite varied for the different source language prepositions. It ranges from 3.2 to 35.8 percentage points, and is generally higher where the baseline is low. The two prepositions that show the highest baseline are *med* and *i*. For these, the most frequent translation is appropriate in more than 80% of the cases. By adding the post-processing filter to these, the accuracy only slightly increases (by 3.2 and 3.7 percentage points respectively). For *på* and *om*, on the other hand, the most frequent translation is appropriate in only 45.3% and 59.3% of the respective cases. Adding the post-processing filter to these dramatically improves the accuracy (by 35.8 and 14.1 percentage points respectively). Intuitively, *med* and *i* are more inclined to be used lexically than are *på* and *om*. This may, in part, explain why the baseline strategy of simply selecting the most frequent translation is so much more effective for the former two prepositions than it is for the latter two.

Summing up the results for all six prepositions, the application of the learnt rule sequences gives an accuracy of 84.9% which corresponds to an increase of 9.4 percentage points as compared to the baseline.

---

[4] The score of a rule is its number of positive instances minus its number of negative instances

[5] The accuracy of a rule is its number of positive instances over its total number of instances.

8. Concluding Remarks

We have reported on an experiment with using transformation-based learning to induce rules to select target language prepositions. Selectional constraints have been sought in the target language context. To avoid loosing control of the source language prepositions, we have used aligned bilingual corpus data only, and induced one rule sequence for each source language preposition.

An evaluation, using the built-in evaluation function in μ-TBL, revealed an accuracy of 84.9% which corresponds to an increase of 9.4 percentage points as compared to the baseline where the most frequent translation is always selected.

It still remains to be investigated how the application of the rule sequences would perform on data output from a real MT-system. The rules are conditioning target words in the context of the prepositions, and the applicability of the rules is thus dependent on the translation of the surrounding words. The effect of this is something which can only be estimated empirically.

9. References

BRANTS, T. (2000). 'TnT – a statistical part-of-speech tagger'. In Proceedings of the 6th Applied NLP Conference (pp. 224-231), Seattle, USA.

BRILL, E. and P. Resnik. (1994). 'A rule-based approach to prepositional phrase attachment disambiguation'. In Proceedings of the 15th conference on Computational Linguistics (pp. 1198-1204), Kyoto, Japan.

BRILL, E. (1995). 'Transformation-based error-driven learning and natural language processing: A case study in part-of-speech tagging'. Computational Linguistics, (21:4):543-566.

BROWN, P., S. Della Pietra, V. Della Pietra, R. Mercer. (1991a). 'A statistical approach to sense disambiguation in machine translation'. In Proceedings of the DARPA Workshop of Speech and Natural Language (pp. 146-151), Pacific Grove, California.

BROWN, P., S. Della Pietra, V. Della Pietra, R. Mercer. (1991b). 'Word Sense Disambiguation using statistical methods'. In Proceedings of the 29th Annual Meeting of the Association for Computational Linguistics (pp. 264-270), Berkeley, California.

DAGAN, I. and A. Itai. (1994). 'Word Sense Disambiguation Using a Second Language Monolingual Corpus'. Computational Linguistics, (20:4):563-596.

EJERHED, E., G. Källgren, O. Wennstedt and M. Åström. (1992). 'Linguistic Annotation System of the Stockholm-Umeå Project'. Technical Report, Department of General Linguistics, University of Umeå.

KANAYAMA, H. (2002). 'An Iterative Algorithm for Translation Acquisition of Adpositions'. In Proceedings of the 9th Conference on Theoretical and Methodological Issues in Machine Translation (pp. 85-95), Keihanna, Japan.

KOEHN, P. (n.d.). 'Europarl: A Multilingual Corpus for Evaluation of Machine Translation'. Draft, Unpublished.

LAGER, T. (1999). 'The μ-TBL System: Logic Programming tools for Transformation-Based Learning'. In Proceedings of the 3rd International Workshop on Computational Natural Language Learning (pp. 33-42), Bergen, Norway.

LAGER, T., N. Zinovjeva. (2001). 'Sense and Deduction: The Power of Peewees Applied to the SENSEVAL-2 Swedish Lexical Sample Task'. In Proceedings of SENSEVAL-2: 2nd International Workshop on Evaluating Word Sense Disambiguation Systems, Toulouse, France.

MANGU, L. and E. Brill. (1997). 'Automatic rule acquisition for spelling correction'. In Proceedings of the 14th International Conference on Machine Learning (pp. 187-194), Nashville, Tennessee.

MARCUS, M., B. Santorini, M.-A. Marcinkiewicz. (1994). 'Building a large annotated corpus of English: The Penn Treebank'. Computational Linguistics, 19(2):313:330.

MEGYESI, B. (2002). 'Data-Driven Syntactic Analysis – Methods and Applications for Swedish'. PhD thesis. Department of Speech, Music and Hearing, KTH, Stockholm, Sweden.

MILLER, K. (1998). 'From above to under: Enabling the Generation of the Correct Preposition from an Interlingual Representation'. In Proceedings of the AMTA/SIG-IL Second Workshop on Interlinguas, Langhorne, Pennsylvania.

MINNEN, G, J. Carroll and D. Pearce. (2001). 'Applied morphological processing of English'. Journal of Natural Language Processing, (7:3):207-223.

OCH, F., H. Ney . (2000). 'Improved Statistical Alignment Models'. In Proceedings of the 38th Annual Meeting of the Association for Computational Linguistics (pp. 440-447), Hong Kong, China.

PALIOURAS, G., V. Karkaletsis, I. Androutsopoulos and C.D. Spyropoulos . (2000). 'Learning Rules for Large- Vocabulary Word Sense Disambiguation: A comparison of Various Classifiers'. In Proceedings of the 2nd International Conference on Natural Language Processing' (pp. 383-394), Patra, Greece.

TSENG, J. L. (2000). 'The Representation and Selection of Prepositions'. PhD Thesis, University of Edinburgh.