Using Noun Class Information to Model Selectional Preferences for Translating Prepositions in SMT

Marion Weller[1,2] wellermn@ims.uni-stuttgart.de
Sabine Schulte im Walde[1] schulte@ims.uni-stuttgart.de
Alexander Fraser[2] fraser@cis.uni-muenchen.de

[1]IMS, Universität Stuttgart, Germany
[2]CIS, Ludwig-Maximilians-Universität München, Germany

Abstract

Translating prepositions is a difficult and under-studied problem in SMT. We present a novel method to improve the translation of prepositions by using noun classes to model their selectional preferences. We compare three variants of noun class information: (i) classes induced from the lexical resource GermaNet or obtained from clusterings based on either (ii) window information or (iii) syntactic features. Furthermore, we experiment with PP rule generalization. While we do not significantly improve over the baseline, our results demonstrate that (i) integrating selectional preferences as rigid class annotation in the parse tree is sub-optimal, and that (ii) clusterings based on window co-occurrence are more robust than syntax-based clusters or GermaNet classes for the task of modeling selectional preferences.

1 Introduction

The translation of prepositions is difficult in SMT: some prepositions convey a meaning (*to sit* UNDER/ON *the table*) while others are merely functional (*to believe* IN *sth.*). Both kinds of prepositions pose a significant challenge to the translation system as they largely depend on target-language-specific restrictions for which there is often not enough contextual information available. Translating prepositions is difficult as SMT systems must choose the correct preposition given the intended meaning of the preposition in the input sentence as well as the target-side context in which the preposition appears.

In English-to-German translation, there are cases in which the target-language does not require a preposition (e.g. *to call* FOR → *fordern*), or in which it is necessary to produce a target-language preposition even though there is no preposition in the input sentence (e.g. *to enter* → *gelangen* IN). The choice of prepositions is typically determined by a governor, such as verbs (*to believe in sth.*) or nouns (e.g. *interest in sth.*). In addition, the preposition depends on the semantic class of nouns that are governed. For example, *to learn from* can lead to two different translations in German: in the case of *to learn from [a person]*, the correct translation is *lernen* VON, whereas *to learn from [the past]* should be translated as *lernen* AUS.

We present a novel method that uses noun class information to model selectional preferences of prepositions. By annotating noun class information into the parse trees used to train an English-German string-to-tree SMT-system, we aim at obtaining more precise translation rules. Instead of allowing any PP in a given rule, the noun class annotation restricts that rule to PPs of a specific semantic class. While this procedure adds semantically fine-grained information, it also leads to a loss of rule generalization. We compensate for this loss by making generic, non-annotated rules accessible for the enriched system, and by generating new PP rules that cannot be derived from the parallel data. The selectional preferences are instantiated by three variants of noun class information:

- nominal concepts induced from the lexical semantic taxonomy *GermaNet*,
- k-Means cluster analyses relying on standard distributional window co-occurrence,
- k-Means cluster analyses relying on syntactic features from dependency-parsed data.

Using noun classes from a lexical resource such as GermaNet allows us to access a conceptually refined form of target-language information. In contrast, by using large target-language corpora as a basis for clustering, we generalize better over contexts (in a "raw" form vs. based on syntactic dependencies) and thus take into account additional target-language information based on very large corpora in a way that goes beyond the potential of an SMT-system that only has access to an n-gram language model.

Even though none of the enriched systems significantly outperforms a baseline without noun class information, our experiments provide insights into the integration of noun classes into a syntactic SMT system regarding (i) the method of annotation and (ii) the resources used. Integrating selectional preferences as rigid annotation in the parse tree is not optimal, as there is no generally applicable optimal level of semantic information. With regard to resources, we found that cluster analyses based on simple window information are better at capturing selectional preferences, with superior performance to both (a) the clusters relying on syntactic features and (b) the classes induced from the high-quality lexical resource GermaNet.

2 Related Work

Translating prepositions is an important problem in machine translation. So far, research has mostly been reported on rule-based systems. Gustavii (2005) uses bilingual features and selectional constraints to correct translations from a rule-based Swedish-English system; she reports a gain in accuracy for prepositions. Naskar and Bandyopadhyay (2006) outline a method to handle prepositions in an English-Bengali MT system: they use WordNet in combination with a bilingual example base for idiomatic PPs, but do not report any evaluation. Agirre et al. (2009) model Basque prepositions and grammatical case using syntactic-semantic features such as subcategorization triples for a rule-based system; they also report a gain in translation accuracy for prepositions. The approach of Shilon et al. (2012) is similar to the work of Agirre et al. (2009); however, Shilon's system has a statistical component for ranking proposed translations, which leads to an improvement in BLEU for a small test set. Furthermore, Zollmann and Vogel (2011) use cluster information in syntactic SMT, although not specifically for translating prepositions.

Huang and Knight (2006) propose methods of relabeling syntax trees to improve statistical syntactic translation. Their annotation aims at making the used tag-set (based on the Penn Treebank) less general, assuming that it often fails to capture relevant grammatical distinctions and contexts that are crucial for translation. They distinguish between *internal* and *external* annotation. In the case of internal annotation, additional information about the node or its relatives that is otherwise not accessible to the respective node is annotated; this type of annotation consists of *lexical* and *tag* information. Their lexicalization strategies include annotating a preposition onto both its parent node (PREP) and its grandparent node (PP), leading to an improvement in BLEU. Other forms of lexicalization consist in annotating information about determiners, auxiliaries and conjunctions. The annotation of tags mainly aims at improving auxiliary and tense errors and is applied to VP-nodes. Furthermore, as external annotation, information about sister nodes and parent nodes, for example, is annotated in order to provide more information about the context of the respective word or phrase. Huang and Knight (2006) report improvements for most of their annotation strategies.

The method presented in this paper is different from the previous approaches as it combines information about subcategorization and noun classes and it is applied using a purely statistical MT system. Furthermore, by annotating noun class information on NPs and PPs, we aim to introduce a semantic level in contrast to the mainly syntactically motivated annotation scheme of Huang and Knight (2006).

3 Obtaining Noun Class Information

Our system relies on noun class information in order to refine hierarchical translation rules such that they incorporate selectional preferences. In this section, we will describe three approaches to obtain noun class information by classifying noun types into semantic classes: (i) assigning nouns to GermaNet classes; (ii) clustering nouns on the basis of window information and (iii) clustering nouns on the basis of syntactic dependency information. Comparing these disjunctive methods should ensure a systematic assessment of integrating selectional preferences.

3.1 Pre-processing

In order to obtain a consistent noun class annotation, we applied two pre-processing steps to the target-language data prior to computing noun classes using the three variants.

In the first step, we attempt to resolve (possibly) inconsistent parsing decisions for word types tagged both as nouns and named entities. Only words recognized as nouns by the high-coverage morphological analyzer SMOR (Schmid et al., 2004) are considered as common nouns. The remaining instances are considered as named entities; they are classified into *organization, location, person* and a category for *rest* (Faruqui and Padó, 2010). Performing this pre-processing ensures that nouns are consistently labeled with the same noun class or named entity category[1]. A second benefit is that only nouns, for which we can expect to have either GermaNet coverage or a sound basis for feature extraction, are considered for clustering; "non-nouns" (such as typos or parse-errors which are often very low-frequency and thus likely to deteriorate clustering performance) are excluded from clustering.

The second pre-processing step consists in compound handling: as German noun compounding is very productive and can lead to sparsity and coverage problems, we applied compound splitting to all nouns using a linguistically-informed compound splitter (Fritzinger and Fraser, 2010), which disambiguates competing SMOR analyses relying on corpus statistics.

After pre-processing, noun class information is first computed for head nouns. Then, in a second step, compounds are added into classes based on their head noun. While this might introduce noise for a small number of non-compositional compounds, we assume that the gain in generalization is more important.

3.2 GermaNet

GermaNet (Hamp and Feldweg, 1997; Kunze, 2000) is a lexical resource for German similar to the English WordNet (Fellbaum, 1998). It is a lexical-semantic taxonomy that groups words of the same concept into synsets. For each head noun, we looked up the GermaNet class for a given hierarchical level, to determine the degree of generalization: GermaNet is graph-structured, and extracting the nouns at different levels results in more or less fine-grained sets of classes. We used noun classes from the levels 2, 3, 4, 5, counting from the top level[2].

[1] Note, however, that our type-based annotation method does not take into account polysemy of regular nouns or names that can also be nouns, e.g. *Zimmermann* ('carpenter'), which is also a common family name.

[2] Words belonging to several synsets (*apple* $\rightarrow$ *plant*|*fruit*) are added to the synset with the lowest GermaNet-internal ID. Nouns not covered by GermaNet (16.357 of 211.360 after compound processing) are assigned to a *rest* class.

Noun		GermaNet (level 5)
Minister	*minister*	human being
Kanzler	*chancellor*	human being
Mehrheit	*majority*	group
Opposition	*opposition*	configuration
Enthebung	*dismissal*	ending/stop

Table 1: A selection of political nouns from an example cluster (window-features) and their GermaNet classes.

3.3 Clustering

For clustering, we used the standard k-Means implementation in R (**?**), with features extracted from the target-side part of the parallel data used to train the SMT system and a large web corpus (ca. 45 M sentences total, cf. section 5.1). Low-frequency nouns (f<5 in the combined corpora) were excluded from clustering, and added to the cluster with the nearest centroid in a post-clustering step. We applied two types of features:

- Content words from a *window* of 10 words to each side of the noun,
- *Syntactically-motivated features* referring to subcategorization criteria:
  1. prepositions governing the target nouns ("P"),
  2. verbs subcategorizing the target nouns ("VO"),
  3. verbs governing the target nouns in a prepositional phrase ("VPN"),
  4. nouns governing the target nouns in a prepositional phrase ("NPN").

We observed that particularly the window-based approach induces "topic-like" clusters, see table 1, where a *politics*-related cluster contains *persons* (*minister, chancellor*) and other terms related to *politics*. In contrast, the classes assigned by GermNet resemble more a generalization over specific noun types as *human beings* (*minister, chancellor*) are grouped together and the remaining terms *majority, opposition* and *dismissal* are each in a separate group. Using syntactic features for clustering, in particular prepositions, aims at better capturing selectional preferences, and thus obtaining classes that provide salient information for the task of modeling the choice of prepositions in SMT; cf. for example Prescher et al. (2000), Erk et al. (2010), Joanis et al. (2008), Schulte im Walde (2006) and **?** for more information.

A major problem consists in finding a number of clusters that provides both (i) a good representation of the nouns and (ii) the optimal level of abstraction for our SMT-system. In our experiments, we varied the cluster sizes and used sets of 10 – 300 clusters.

4 Using Noun Class Information in SMT

This section presents the basic enriched system and its variants extended with non-annotated baseline rules and new PP rules. In all experiments, a preposition is annotated on both its parent node (PREP) and its grandparent node (PP), as suggested by Huang and Knight (2006).

4.1 Annotating Rules with Noun Classes

Figure 1 illustrates how the target-language parse trees are annotated with noun class information by introducing indices for NP and PP nodes, nouns and prepositions. In this example[3],

[3]We work with a string-to-tree system: annotations on the English side of the rules are given only for better readability.

```
<tree ="s">
  <tree="adjd"> wirtschaftlich</tree>
  <tree="vafin-haben"> hat </tree>
  <tree="np-LOC">
    <tree="ne-LOC"> malaysia </tree>
  </tree>
  <tree="vp">
    <tree="pp-von-167">
      <tree="prep-von-167"> von </tree>
      <tree="pposat"> seinen </tree>
      <tree="nn-167"> nachbarn </tree>
    </tree>
    <tree="vvpp"> gelernt </tree>
  </tree>
</tree>
```

economically, Malaysia has **learned from its neighbors**.

```
<tree ="s">
  <tree="kous"> dass </tree>
  <tree="np-180">
    <tree="art"> die </tree>
    <tree="nn-180"> amerikaner </tree>
  </tree>
  <tree="vp">
    <tree="pp-aus-291">
      <tree="prep-aus-291"> aus </tree>
      <tree="art"> der </tree>
      <tree="nn-291"> vergangenheit </tree>
    </tree>
    <tree="vvpp"> gelernt </tree>
  </tree>
  <tree="vafin-haben"> hätten </tree>
</tree>
```

that the Americans had **learned from their past**.

Figure 1: Example for annotating parsed data with noun class information.

the noun class information serves to create two variants for the translation of `learned [from NOUN]`$_{PP}$, namely

`VP` $\rightarrow$ `PP-von-167 gelernt`

and

`VP` $\rightarrow$ `PP-aus-291 gelernt`,

indicating that nouns of the classes 167 (*person*) and 291 (*abstract concept*) represent appropriate fillers for the respective PPs subcategorized by the verb *lernen* ('*to learn*'), headed by the prepositions "von" and "aus", respectively.

4.2 Adding Non-annotated NP+PP Rules

Noun class annotation on NP and PP nodes might lead to overly specific rules, resulting in a loss of rule generalization in comparison to the baseline. We thus added baseline rules (rules without cluster annotation) to the enriched rules (we will call this system "BL" in the experimental section). Rules derived from source-target pairs that occurred with $f \leq 5$ are likely to be random and not useful for selection preferences, so we removed them, leaving only the non-annotated rules (system "BL+cutoff"). Alternatively, we only kept baseline rules with a higher translation probability than the respective annotated rules, thus favoring annotated rules. Rules such as `to buy nn1/nn2/nn3/...` are replaced with `to buy nn` (system "BL-subst").

4.3 Generating New PP Rules

In addition to the problem that annotated rules can be too specific, not all potentially necessary rules can be obtained from the parallel data. New PP rules are generated by duplicating existing annotated PP rules in which prepositions are substituted. This creates new rules that are not accessible to the baseline and aims at providing the full possible set of rules containing functional prepositions, i.e. prepositions conveying no or only little meaning. Assuming that functional or subcategorized prepositions are the most difficult to translate, the prepositions for which to generate new rules rely on the set of subcategorized prepositions in a subcategorization lexicon (Eckle (1999)). This set comprises 17 prepositions: *an, auf, aus, bei, durch, für, in, mit, nach, über, um, unter, von, vor, wegen, zu, zwischen*.

source-side: `learn [X`$_{PP}$`] , [X`$_S$`]`	
original rule (target-side)	**prob.**
`VP → [pp-von-166] lernen , [s]`	1
new PP rules (target-side)	**prob.**
`VP → [pp-aus-166] lernen , [s]`	0.159
`VP → [pp-für-166] lernen , [s]`	0.021
`VP → [pp-in-166] lernen , [s]`	0.126
`VP → [pp-mit-166] lernen , [s]`	0.021
`VP → [pp-von-166] lernen , [s]`	0.336
`VP → [pp-über-166] lernen , [s]`	0.336

Table 2: Newly generated PP translation rules.

prep-noun-verb tuple		**freq**
aus nn-166 lernen	*to learn from nn-166*	38
für nn-166 lernen	*to learn for nn-166*	5
in nn-166 lernen	*to learn in nn-166*	30
mit nn-166 lernen	*to learn with nn-166*	5
von nn-166 lernen	*to learn from nn-166*	80
über nn-166 lernen	*to learn about nn-166*	80

Table 3: Subcategorization tuples induced from large monolingual data.

Table 2 shows how the target-side of the original (annotated) rule is multiplied into six new rules containing the prepositions observed in combination with the verb *lernen* and nouns of class 166. The translation probabilities are derived from co-occurrence frequencies in the combined web and target-side part of the parallel corpus (cf. table 3), such as tuples of the form *n-prep-n*, *prep-n-verb*, etc. Only PP-nodes or PREP-nodes are modified, the rest of the rule (other nodes, terminal symbols and the source-side) remains the same. To keep the amount of generated rules manageable, we used a threshold of $f \geq 5$ to select the rules for which to generate new PP rules and only kept generated rules with a translation probability of $p \geq 0.001$ ("new rules"). Finally, we added both baseline and new rules ("BL+new").

5 Experiments and Results

We used a morphology-aware English-German translation system that first translates into a lemmatized representation, and then generates inflected forms based on morphological features predicted with a sequence model (e.g. Toutanova et al. (2008), Fraser et al. (2012)). This reduces morphological complexity of nominal phrases, and allows in particular to handle portmanteaus (combination of preposition and article: *zur=zu+der*: *to the*) which are split in pre-processing and merged in a post-processing step. Thus, during translation, prepositions occurring as portmanteaus are represented in the same way as non-portmanteau prepositions.

Table 4 illustrates the processing steps. The lemmatized representation (first column) contains feature markup on nouns for the features *number* and *gender*, which are considered part of the stem. The information about *gender* is obtained from a morphological resource and typically does not vary for a given noun, whereas *number* is indirectly determined by the source-side; as-

SMT output + stem markup	**predicted features**	**generated forms**	**gloss**
konzentriert[VVFIN]	–	konzentriert	*concentrates*
sich[PRO]	–	sich	*itself (refl. pron.)*
auf[PREP]	–	auf	*on*
Bemühung<**Fem**><**Pl**>[NN]	**Fem**.Acc.**Pl**.St	Bemühungen	*efforts*
zu[PREP]	–	zu ⇒ **zur**	*on*
die<+ART>[ARTdef]	Fem.Dat.Sg.St	der	*the*
Verringerung<**Fem**><**Sg**>[NN]	**Fem**.Dat.**Sg**.Wk	Verringerung	*reduction*
die<+ART>[ARTdef]	Fem.Gen.Pl.St	der	*of*
Treibhausgasemission<**Fem**><**Pl**>[NN]	**Fem**.Gen.**Pl**.Wk	Treibhausgas-emissionen	*greenhouse gas emissions*

Table 4: Processing steps for the input sentence "... *focuses on efforts to cut greenhouse gas emissions* ..." including portmanteau merging as last post-processing step.

suming that the number of a source-side noun is preserved during translation.

To predict morphological features, we trained a sequence model for the features *number*, *gender*, *case* and *strong/weak inflection*. Each model has access to stems, POS-tags and the feature to be modelled within a window of four positions to the right and the left of the current position. The stem-markup is part of the input to the feature prediction step and is basically propagated over the rest of the phrase, whereas the features *case* and *strong/weak inflection* are predicted solely based on context information, i.e. adjacent tags and stems (second column).

Based on the predicted morphological features and the lemma, inflected forms can be generated using a morphological resource (third column). Finally, after generating inflected forms, split instances of portmanteau prepositions are merged relying on a simple set of rules[4] as illustrated in the example (*zu+der* → *zur*) in the third column of table 4.

5.1 Data

We used 1.5 M sentences of parallel data (Europarl and news data from the 2009 WMT shared task), with the target-side part as language model data, to train a string-to-tree Moses system with GHKM extraction (Galley et al., 2004; Williams and Koehn, 2012). The tuning/test sets consist of 1025/1026 news sentences (from the 2009 WMT shared task). The German data was parsed with BitPar (Schmid, 2004). For generating inflected forms, we used the morphological tool SMOR (Schmid et al., 2004). For predicting the morphological features *number*, *gender*, *case* and *strong/weak inflection*, we trained one CRF for each of the four morphological features using the Wapiti toolkit (Lavergne et al., 2010).

The tuples for modelling translation probabilities for rule generation and the context vectors for clustering were obtained from a combination of the web corpus *SdeWaC* (44M sentences, Faaß and Eckart (2013)) and the German part of the parallel data.

5.2 Results

Table 5 presents the results of the systems enriched with noun class information; none of the systems is significantly better than the baseline without semantic class information. Interestingly, the window-based cluster systems are better than the systems using GermaNet or syntactic fea-

[4]In contrast to Romance languages, where the merging of portmanteaus (e.g. *à+le=au*) is mandatory, the merging of German portmanteaus is not always necessary. However, preliminary experiments indicated that merging whenever possible is a good strategy.

System	BLEU
Baseline	13.95
GermaNet-2 (25)	13.93
GermaNet-3 (79)	13.77
GermaNet-4 (175)	13.67
GermaNet-5 (392)	13.67

System	BLEU
Window10	14.01
Window50	14.18
Window75	13.69
Window100	14.13
Window300	13.71

Syntactic features	**P**	**VO**	**VPN**	**NPN**
100 classes	13.85	13.85	13.79	13.71
50 classes	13.84	14.06	14.06	13.91

Table 5: Results for different annotation settings: GermaNet and clusterings based on window information or syntactic features; the scores are averaged over two tuning runs. The numbers in brackets for GermaNet indicate the number of classes and the numbers 2,3,4,5 denote the respective level.

System	**BL**	**BL cutoff**	**BL subst**	**new rules**	**BL new**
Window50	13.95	13.99	14.04	14.11	13.98
Window75	14.16	13.96	14.07	13.66	14.01
Window100	14.01	13.94	13.96	14.14	14.02

Table 6: System variants with non-annotated rules and new PP rules.

tures. While GermaNet is a high-quality resource, it tends to suffer from coverage problems and is too fine-grained (for example, the word *chancellor* is assigned to 2 classes at level 5: *organism* and *living being*, which is a distinction that is not needed in our application). On the other hand, the syntactic features are more sparse than window-based features. This is due to the simple fact that we can nearly always extract content words within a window for a given noun, but the extraction of syntactic features is more restrictive and thus, features can only be extracted in the respective syntactic constellation. The window clusters thus seem to provide the most robust representation of selectional preferences. The number of classes does not seem to have a strong overall influence, even though there is a tendency for less classes being favorable.

For three systems (Window50/75/100), we added non-annotated rules ("BL", "BL-cutoff", "BL-subst"), new PP rules ("new rules") and a combination of new and non-annotated rules ("BL+new"), cf. table 6. While there is a moderate improvement for Window75, one of the worst systems in table 5, there is no further gain for the other two systems.

When analyzing the enriched systems' output, we noticed that on average, more and shorter translation rules than in the baseline systems were used. For example, the enriched systems Window50/75 use on average 11.99/11.62 glue rules per sentence, whereas the baseline system only uses 7.10 glue rules on average. Similarly, the average rule length (here: the length of the target-side of a rule) decreases from 2.19 (baseline) to 1.91/1.92 for the window systems. The average sentence length is stable over these three systems, varying between 25.3 and 25.5 words. Assuming that the use of a low amount of glue rules and long translation rules is preferable, we consider this an indicator of a general problem with the enriched rules: longer and more specific rules in the enriched system do not match anymore and are thus replaced by a combination of shorter rules, resulting in a loss of the context provided by a single longer rule. This contradicts our initial objective of annotating noun class information to add new, generalized information about nouns in order to provide a better basis to model selectional preferences

EN	more than $ 100 billion will **enter the monetary markets** by means of public sales.
BL	mehr als 100 Milliarden Dollar wird **die Geldmärkte** durch öffentlichen Verkauf **gelangen**. *more than 100 billion dollar will* **get** ∅ **money markets** *by means of public sale.*
W	mehr als 100 Milliarden Dollar **auf die Geldmärkte gelangen** wird durch den öffentlichen Verkauf. *more than 100 billion dollar* **get on the money markets** *by means of the public sale.*

EN	the charge that she **concentrated** too much **on foreign affairs** , ...
BL	der Vorwurf , dass sie **auswärtige Angelegenheiten** zu stark **konzentriert** ist , ... *the charge, that she* ∅ **foreign affairs** *too strong* **concentrated** *is, ...*
W	der Vorwurf , dass sie zu sehr **auf die auswärtigen Angelegenheiten konzentriert** , ... *the charge, that she too much* **on the foreign affairs concentrates***, ...*

EN	one of the local residents even classified the quarrels with eastern european immigrants as a fight **for** survival.
BL	eine der Anwohner selbst ein Kampf **für** das Überleben der Streitigkeiten mit osteuropäischen Migranten eingestuft . *one of the residents even the fight for the survival of-the quarrels with eastern european immigrants classified.*
W	eine der Anwohner sogar eingestuft die Streitigkeiten mit osteuropäischen Einwanderern wie ein Kampf **ums** Überleben. *one of the residents even classified the quarrels with eastern european immigrants like a fight for survival*

Table 7: Examples for better translation of prepositions (BL=Baseline, W=Window50).

of prepositions. Thus, introducing noun classes as a new form of information by the means of parse-tree annotation comes at the cost of losing basic context information as rules spanning over larger chunks are often not available anymore.

5.3 Examples of Improved Translations

Table 7 gives three examples of improvements obtained with the enriched system: in the first sentence, the translation of *enter* → *gelangen* requires the preposition *auf* (*to get on*), which is correctly produced by the enriched system. Note that it is also possible to translate the phrase *enter the money markets* without using a preposition in German, for example with the verb *erreichen*+DIRECT OBJECT (*to reach the money markets*).

In the second sentence, the preposition for the translation of *concentrate on* is missing in the baseline, but is correctly produced by the enriched system. In the third sentence, the phrase *Kampf für das Überleben* (*fight for the survival*) is somewhat understandable, but the preposition *ums* (portmanteau: *um+das*) in the enriched system is a much better choice.

6 Evaluation and Discussion

In this section, we present a more in-depth evaluation in form of assessing the translation quality of prepositions. Furthermore, we illustrate typical problems encountered when translating prepositions, but also show examples for improved sentences. Finally, we discuss why the noun class annotation did not lead to more improvement.

	for	**on**	**in**	**at**
Baseline	40	24	81	15
Window50	42	25	85	18
Total	59	48	110	30

Table 8: Correctly translated English prepositions.

6.1 Translation quality of prepositions

In addition to applying BLEU, we manually evaluated the translation quality of English prepositions, using a set of sentences (5–20 words long) containing the prepositions *for, on, in*, or *at*. This test set also includes sentences where a translation of the English preposition as a "null" preposition is necessary or possible, as illustrated in the following examples:

(1a) that lead to a knock-on fall **in** exports to western europe

(1b) das führt zu einem erheblichen Rückgang **der** Exporte nach Westeuropa

that lead to a considerable fall **the**$_{GEN}$*(=of the) exports to western europe*

(2a) ... has again commented **on** the problem of global warming

(2b) ... hat erneut $\emptyset$ das Problem der globalen Erwärmung kommentiert

In (1) , the preposition *in* can be expressed in form of a genitive modification (*der Exporte*), but a translation as preposition is also possible. In (2), it is not possible to translate the preposition *on* when using the verb *kommentieren* which requires a direct object. However, with the verb *sich äußern* as translation of *to comment*, a preposition (*zu*) is required.

The fact that often several translation variants (e.g. depending on the choice of the verb) are possible makes it difficult to directly compare the systems' output to a reference translation. We considered a preposition to be correctly translated if the produced PP or NP is an acceptable translation in the German sentence. Table 8 shows that the enriched system is slightly better than the baseline system, but overall there is only a small difference.

In general, we found it difficult to observe a systematic behaviour or "pattern" of (types of) prepositions or contexts that are handled better or worse in the enriched system in comparison to the baseline system. However, we noticed that there is a type of prepositions that seems to be especially hard to translate, namely prepositions with a predominantly literal meaning occurring in an infrequent subcategorized context. These are often mistranslated, in both our baseline and enriched systems, as illustrated in the following example:

(3a) for example , germany has been criticized **for** passivity

(3b) beispielsweise hat Deutschland ***für** Passivität kritisiert worden

for example, Germany has ***for** *passivity criticized been*

(3c) **wegen** Passivität wurde zum Beispiel Deutschland kritisiert

The preposition *for* is often used literally and thus can be translated in a straightforward manner, e.g. with the preposition *für*. In this subcategorized context (*criticized for*) however, it expresses a cause, which makes *für* a totally inappropriate translation, cf. (3b). In contrast, *wegen* (*because of*) is the correct translation, as can be seen in the reference translation (3c). We noticed that similar constructions such as "*detain* FOR *corruption*" (WEGEN *Korruption verhaften*) or "*look* FOR *sth.*" (NACH *etwas suchen*) seem to be prone to the same error.

These examples provide insight into the complexity of the task of translating prepositions. Depending on the respective context of a PP, different factors such as the relation of being a merely functional preposition (i.e. subcategorized) vs. conveying a meaning, as well as the class of the involved noun seem to play roles of varying importance. However, with our rather inflexible annotation method we are not able to act in a context-dependent manner, but always provide the same type of information at the same level of granularity.

6.2 Conclusion

We started with the hypothesis that noun class information is useful to model selectional preferences in preposition translation rules. However, annotating semantic class information on NP/PP nodes of the parse trees in a string-to-tree system amounts to a hard constraint and our experiments indicate that this form of annotation leads to overly specific rules. We tried to compensate for this by making the non-annotated rules available and by adding new PP rules synthesized from monolingual data. However, previous work, such as e.g. the work of Marton and Resnik (2008), has shown that soft constraints often work better than hard constraints. It might therefore make sense to model selectional preferences through the use of feature functions which reward good choices, rather than markup in the string-to-tree grammar, but this would require extensive changes to the model and decoder.

Another problem with our approach is that there is no generally applicable optimal level of selectional preferences. This is in line with semantic research on selectional preferences as verb subcategorization features (Schulte im Walde, 2006; Joanis et al., 2008): across subcategorizing words, it is difficult to identify a generally acceptable semantic level of generalization in lexical resources. Because of this, the parse tree annotation is not flexible enough to take into account the varying needs of different contexts, as it always leads to rules of the same degree of specificity, and therefore cannot adapt to the respective contexts.

With regard to resources, we found that none of the variants we considered was able to obtain noun class information that is optimal: WordNets in general are known to be very fine-grained and contain many ambiguities, making it difficult to derive generally applicable noun groups (Navigli, 2006; Palmer et al., 2007). In contrast, window clusters might not contain the appropriate selectional preference information as they resemble topic clusters rather than a generalization over specific noun types. As opposed to the unstructured information used for the window clustering, the syntactic dependencies constitute the type of information that is needed to determine a valid preposition for a given context, i.e. the governing verb/noun or the noun in the PP. Thus, clusters learned from syntactic features were expected to better capture selectional preferences. However, these clusters failed to lead to improvements and had worse performance than the window clustering.

This work thoroughly explored different methods to obtain noun class information (exploiting distributional and resource-based information), but found that none of these variants is optimal. While each of these strategies has some advantages (e.g. either high-quality or high coverage), they also suffer from weaknesses (low coverage or too fine-grained) that could be (at least partially) addressed by combination with another method. For example, GermaNet could be used to provide a high-quality initial set of noun classes that is then expanded relying on distributional information providing a wide-range coverage. Combining the advantages of the resources we considered could lead to a more promising strategy to obtain classes providing salient information on selectional preferences and constitutes a challenging task for future work.

Acknowledgements

This work was funded by the DFG Research Projects *Distributional Approaches to Semantic Relatedness* and *Models of Morphosyntax for Statistical Machine Translation – Phase 2* and the *DFG Heisenberg Fellowship SCHU-2580/1-1*. We would like to thank several anonymous reviewers for their helpful comments.

References

Agirre, E., Atutxa, A., Labaka, G., Lersundi, M., Mayor, A., and Sarasola, K. (2009). Use of Rich Linguistic Information to Translate Prepositions and Grammatical Cases to Basque. In *Proceedings of the 13th Annual Conference of the EAMT*.

Eckle, J. (1999). *Linguistisches Wissen zur automatischen Lexikon-Akquisition aus deutschen Textcorpora*. PhD thesis, Institut für Maschinelle Sprachverarbeitung, Universität Stuttgart.

Erk, K., Padó, S., and Padó, U. (2010). A Flexible, Corpus-Driven Model of Regular and Inverse Selectional Preferences. *Computational Linguistics*, 36(4):723–763.

Faaß, G. and Eckart, K. (2013). SdeWaC – a Corpus of Parsable Sentences from the Web. In *Proceedings of the International Conference of the German Society for Computational Linguistics and Language Technology*, Darmstadt, Germany.

Faruqui, M. and Padó, S. (2010). Training and Evaluating a German Named Entity Recognizer with Semantic Generalization. In *Proceedings of KONVENS 2010*, Saarbrücken, Germany.

Fellbaum, C., editor (1998). *WordNet – An Electronic Lexical Database*. Language, Speech, and Communication. MIT Press, Cambridge, MA.

Fraser, A., Weller, M., Cahill, A., and Cap, F. (2012). Modeling Inflection and Word-Formation in SMT. In *Proceedings of the the European Chapter of the Association for Computational Linguistics (EACL)*, Avignon, France.

Fritzinger, F. and Fraser, A. (2010). How to avoid burning ducks: Combining linguistic analysis and corpus statistics for German compound processing. In *Proceedings des ACL 2010 Joint Fifth Workshop on Statistical Machine Translation and Metrics MATR*.

Galley, M., Hopkins, M., Knight, K., and Marcu, D. (2004). What's in a Translation Rule? In *Proceedings of NAACL-HLT 2004*.

Gustavii, E. (2005). Target-Language Preposition Selection - an Experiment with Transformation-Based Learning and Aligned Bilingual Data. In *Proceedings of EAMT 2005*.

Hamp, B. and Feldweg, H. (1997). GermaNet - a Lexical-Semantic Net for German. In *Proceedings of ACL workshop Automatic Information Extraction and Building of Lexical Semantic Resources for NLP Applications*.

Huang, B. and Knight, K. (2006). Relabeling Syntax Trees to Improve Syntax-Based Machine Translation Quality. In *Proceedings of the Human Language Technology Conference of the North American Chapter of the ACL*.

Joanis, E., Stevenson, S., and James, D. (2008). A General Feature Space for Automatic Verb Classification. *Natural Language Engineering*, 14(3):337–367.

Kunze, C. (2000). Extension and Use of GermaNet, a Lexical-Semantic Database. In *Proceedings of the 2nd International Conference on Language Resources and Evaluation*, pages 999–1002, Athens, Greece.

Lavergne, T., Cappé, O., and Yvon, F. (2010). Practical very large scale CRFs. In *Proceedings the 48th Annual Meeting of the Association for Computational Linguistics (ACL)*, pages 504–513. Association for Computational Linguistics.

Marton, Y. and Resnik, P. (2008). Soft syntactic constraints for hierarchical phrased-based translation. In *Proceedings of ACL-08: HLT*, pages 1003–1011, Columbus, Ohio. Association for Computational Linguistics.

Naskar, S. K. and Bandyopadhyay, S. (2006). Handling of Prepositions in English to Bengali Machine Translation. In *Proceedings of the Third ACL-SIGSEM Workshop on Prepositions*.

Navigli, R. (2006). Meaningful clustering of senses helps boost word sense disambiguation performance. In *Proceedings of the 21st International Conference on Computational Linguistics and 44th Annual Meeting of the Association for Computational Linguistics*, Sydney, Australia.

Palmer, M., Dang, H., and Fellbaum, C. (2007). Making Fine-grained and Coarse-grained Sense Distinctions, both Manually and Automatically. *Journal of Natural Language Engineering*, 13(2):137–163.

Prescher, D., Riezler, S., and Rooth, M. (2000). Using a Probabilistic Class-Based Lexicon for Lexical Ambiguity Resolution. In *In Proceedings of the 18th International Conference on Computational Linguistics (COLING 2000)*, Saarbrücken, Germany.

Schmid, H. (2004). Efficient Parsing of Highly Ambiguous Context-Free Grammars with Bit Vectors. In *Proceedings of the International Conference on Computational Linguistics*.

Schmid, H., Fitschen, A., and Heid, U. (2004). SMOR: a German Computational Morphology Covering Derivation, Composition, and Inflection. In *Proceedings LREC 2004*.

Schulte im Walde, S. (2006). Experiments on the automatic induction of German semantic verb classes. *Computational Linguistics*, 32(2):159–194.

Shilon, R., Fadida, H., and Wintner, S. (2012). Incorporating Linguistic Knowledge in Statistical Machine Translation: Translating Prepositions. In *Proceedings of the Workshop on Innovative Hybrid Approaches to the Processing of Textual Data, EACL 2012*.

Toutanova, K., Suzuki, H., and Ruopp, A. (2008). Applying Morphology Generation Models to Machine Translation. In *Proceedings of ACL08-HLT*.

Williams, P. and Koehn, P. (2012). GHKM-Rule Extraction and Scope-3 Parsing in Moses. In *Proceedings of the 7th Workshop on Statistical Machine Translation, ACL*.

Zollmann, A. and Vogel, S. (2011). A Word-Class Approach to Labeling PSCFG Rules for Machine Translation. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics*.