Target-side Generation of Prepositions for SMT

Marion Weller[1,2], **Alexander Fraser**[2], **Sabine Schulte im Walde**[1]

[1] IMS, University of Stuttgart – (wellermn|schulte)@ims.uni-stuttgart.de
[2] CIS, Ludwig-Maximilian University of Munich – fraser@cis.uni-muenchen.de

Abstract

We present a translation system that models the selection of prepositions in a target-side generation component. This novel approach allows the modeling of all subcategorized elements of a verb as either NPs or PPs according to target-side requirements relying on source and target side features. The BLEU scores are encouraging, but fail to surpass the baseline. We additionally evaluate the preposition accuracy for a carefully selected subset and discuss how typical problems of translating prepositions can be modeled with our method.

1 Introduction

The translation of prepositions is a difficult task for machine translation; a preposition must convey the source-side meaning while also meeting target-side constraints. This requires information that is not always directly accessible in an SMT system. Prepositions are typically determined by governors, such as verbs (*to believe in sth.*) or nouns (*interest in sth.*). Functional prepositions tend to convey little meaning and mostly depend on target-side restrictions, whereas content-bearing prepositions are largely determined by the source-side, but may also be subject to target-side requirements, as in the following example: *go <u>to</u> the cinema/<u>to</u> the beach* → *<u>ins</u> Kino/<u>an</u> den Strand gehen*.

In this paper, we treat prepositions as a target-side generation problem and move the selection of prepositions out of the translation system into a post-processing component. During translation, we use an abstract representation of prepositions as a place-holder that serves as a basis for the generation of prepositions in the post-processing step. In this step, all subcategorized elements of a verb are considered and allotted to their respective functions – as PPs with an overt preposition, but also as NPs with an "empty" preposition, e.g. *to call <u>for</u> sth.* → *<u>∅</u> etw. erfordern*. In a standard SMT system, subcategorization is difficult to capture in the language model or by the translation rules if the verb and its subcategorized elements are not adjacent.

In the following, we outline a method to handle prepositions with a target-side generation model in an English-German morphology-aware SMT system. We study two aspects: (i) features for a meaningful abstract representation of prepositions and (ii) how to predict prepositions in the translation output using a combination of source and target-side information. In addition, we compare prepositions in the machine translation output with those in the reference translation for a selected subset. Finally, we discuss examples illustrating typical problems of translating prepositions.

2 Related Work

Most research on translating prepositions has been reported for rule-based systems. Naskar and Bandyopadhyay (2006) outline a method to handle prepositions in an English-Bengali MT system using WordNet and an example base for idiomatic PPs. Gustavii (2005) uses bilingual features and selectional constraints to correct translations in a Swedish-English system. Agirre et al. (2009) model Basque prepositions and grammatical case using syntactic-semantic features such as subcategorization triples for a rule-based system which leads to an improved translation quality for prepositions. Shilon et al. (2012) extend this approach

input	lemmatized SMT output	prep	morph. feat.	inflected	gloss
∅ ⟶	PREP	∅-Acc	–		
what	welch<PWAT>	Acc	Acc.Fem.Sg.Wk	welche	which
role	Rolle<+NN><Fem><Sg>	Acc	Acc.Fem.Sg.Wk	Rolle	role
∅ ⟶	PREP	∅-Nom	–		
the	die<+ART><Def>	Nom	Nom.Masc.Sg.St	der	the
giant	riesig<ADJ>	Nom	Nom.Masc.Sg.Wk	riesige	giant
planet	Planet<+NN><Masc><Sg>	Nom	Nom.Masc.Sg.Wk	Planet	planet
has	gespielt<VVPP>	–	–	gespielt	played
played	hat<VAFIN>	–	–	hat	has
in ⟶	PREP	bei-Dat	–	bei	for
the	die<+ART><Def>	Dat	Dat.Fem.Sg.St	der	the
development	Entwicklung<+NN><Fem><Sg>	Dat	Dat.Fem.Sg.Wk	Entwicklung	development
of ⟶	PREP	∅-Gen	–		
the	die<+ART><Def>	Gen	Gen.Neut.Sg.St	des	of-the
solar system	Sonnensystem<+NN><Neut><Sg>	Gen	Gen.Neut.Sg.Wk	Sonnensystems	solar system

Figure 1: Prediction of prepositions, morphological features and generation of inflected forms for the lemmatized SMT output. German cases: Acc-Accusative, Nom-Nominative, Dat-Dative, Gen-Genitive.

with a statistical component for ranking translations. Weller et al. (2014) use noun class information as tree labels in syntactic SMT to model selectional preferences of prepositions. The presented work is similar to that of Agirre et al. (2009), but is applied to a fully statistical MT system. The main difference is that Agirre et al. (2009) use linguistic information to select appropriate translation rules, whereas we generate prepositions in a post-processing step.

A related task to generating prepositions is the generation of determiners, which are problematic when translating from languages without definiteness morphemes, e.g. Czech or Russian. Tsvetkov et al. (2013) create synthetic translation options to augment a standard phrase-table. They use a classifier trained on local contextual features to predict whether to generate or remove determiners for the target-side of translation rules. Another related task is error correction of second language learners, e.g. Rozovskaya and Roth (2013), which also comprises the correction of prepositions.

In addition to the standard evaluation metric BLEU, we evaluate the accuracy of prepositions in cases where the governing verb and governed noun in the translation output match with the reference translation. Conceptually, this is loosely related to semantically focused metrics (e.g. MEANT, Lo and Wu (2011)), as we go beyond a "flat" n-gram matching but evaluate a meaningful entity, in our case a preposition-noun-verb triple.

3 Methodology

Our approach is integrated into an English-German morphology-aware SMT system which first translates into a lemmatized representation with a component to generate fully inflected forms in a second step, an approach similar to the work by Toutanova et al. (2008) and Fraser et al. (2012). The inflection requires the modeling of the grammatical *case* of noun phrases (among other features), which corresponds to determining the syntactic function[1]. Weller et al. (2013) describe modeling *case* in SMT; we want to treat all subcategorized elements of a verb in one step and extend their setup to cover the prediction of prepositions in both PP and NPs (i.e., the "empty" preposition).

3.1 Translation and Prediction Steps

To build the translation model, we use an abstract target-language representation in which nouns, adjectives and articles are lemmatized and prepositions are substituted with place-holders. Additionally, "empty" place-holder prepositions are inserted at the beginning of noun phrases. To obtain a symmetric data structure, place-holders for "empty" prepositions are also added to source NPs.

When generating surface forms for the translation output, a phrase containing a place-holder can be realized as a noun phrase (with an "empty" preposition) or as an overt prepositional phrase (by generating the preposition's surface form).

Figure 1 illustrates the process: for the English input with extra null-prepositions (column 1), the SMT system outputs a lemmatized representation with place-holder prepositions (column 2). In a first step, prepositions and *case* for the SMT output are predicted (column 3). Then, the three remaining inflection-relevant morphological features *number*, *gender* and *strong/weak* are predicted on "regular" sentences without place-holders, given

[1]The subject usually is in *nominative* case and direct/indirect objects are *accusative/dative*.

the prepositions from the previous step (column 4). In the last step, fully inflected forms[2] are produced based on features and lemmas (column 5). As the inflected forms are generated at the end of the pipeline, portmanteau prepositions, i.e. prepositions merged with an article in certain conditions, such as *zu+dem=zum* (*to+the*), are easily handled.

Due to the lemmatized representation, all subcategorized elements of a verb are available in an abstract form and can be allotted to their respective functions (subject, object, PPs) and be inflected accordingly. Furthermore, the generation of (functional) prepositions is independent of structural mismatches of source and target side: for example, as translation of *to pay attention <u>to</u> sth.*, both *<u>auf</u> etw. achten* and *<u>∅</u> etw. beachten* are possible, but require a different realization of the place-holder (∅ vs. overt preposition).

For the prediction of prepositions, we combine source and target-side features into a first-order linear chain CRF which provides a flexible framework to make use of different knowledge sources. We use distributional information about subcategorization preferences to model functional prepositions, whereas source-side features (such as the aligned word) tend to be more important for predicting prepositions conveying content. These features address both functional and content-bearing prepositions, but are designed to not require an explicit distinction between the two categories because the model is optimized on the relevant features for each context during training.

During the generation step, the relevant information (such as governing verb/noun and subcategorization preferences) is presented in a refined form, as opposed to the limited information available in a standard SMT system (such as immediate context in a translation rule or language model). It is thus able to bridge large distances between the verb and its subcategorized elements.

4 Abstract Representation of Prepositions

In addition to providing a means to handle subcategorized elements by target-side generation, one objective of the reduced representation of prepositions is to obtain a more general SMT system with a generally improved translation performance. Our experiments will show, however, that replacing prepositions by simple place-holders decreases the translation quality. The effect that a simplified SMT system loses discriminative power has also been observed by e.g. Toutanova et al. (2008) who found that keeping morphological information during translation can be preferable to removing it from the system despite the problem of increased data sparseness. We will thus evaluate systems with varying levels of information annotated to the place-holders (cf. section 6.2).

As an extension to the basic approach with plain place-holders, we experiment with enriching the place-holders such that they contain more relevant information and represent the content of a preposition while still being abstract. To this end, we enrich the place-holders with syntactically motivated features. For example, the representation can be enriched by annotating the place-holder with the grammatical case of the preposition it represents: for overt prepositions, case is often an indicator of the content (such as direction/location), whereas for empty prepositions (NPs), case indicates the syntactic function. As extension, we mark whether a place-holder is governed by a noun or a verb.

Furthermore, we take into account whether a preposition is functional or conveys content: based on a subcategorization lexicon (Eckle, 1999), we decide whether a place-holder in a given context is subcategorized or not. This idea is extended to a system containing both place-holder and normal prepositions: assuming that merely functional prepositions contribute less in terms of meaning, these are replaced by an abstract representation (case and type of governor), whereas for all non-functional prepositions, the actual preposition with annotation (case and type of governor) are kept.

5 Predicting Prepositions

In this section, we explain the features used to predict the values of the place-holder prepositions and evaluate the prediction quality on clean data.

5.1 Features for Predicting Prepositions

Table 1 illustrates the features for predicting prepositions: in addition to target-side context in the form of adjacent lemmas and POS-tags (5 words left/right), we combine three types of features: (1) source-side features, (2) projected source-side information and (3) target-side subcategorization frames. The source-side information consists of

- the word aligned to the place-holder preposition: a source-side overt or empty preposition

[2]We only generate inflected forms for NPs/PPs (nouns, adjectives, determiners); verbs are inflected throughout the system.

lemma	gloss	source-side prp	source-side func,noun	source-side g.verb	projected source-side noun	projected source-side g.verb	target-side subcat	label
aber	*but*	–	–	–	–	–	–	-
PRP	*PRP*	∅	subj, we	endure	wir	leiden	∅-Nom:5 ∅-Acc:0 *unter-Dat*:4	∅-Nom
wir	*we*	–	–	–	–	–	–	Nom
leiden	*suffer*	–	–	–	–	–	–	-
...	...	...	...	...	...	...	...	...
auch	*too*	–	–	–	–	–	–	-
PRP	*PRP*	∅	obj, effect	endure	Treibhauseffekt	leiden	∅-Nom:5 ∅-Acc:0 *unter-Dat*:4	unter-Dat
die	*the*	–	–	–	–	–	–	Dat
Treibhaus effekt	*greenhouse effect*	–	–	–	–	–	–	Dat

Table 1: Prediction features in the training data. Source-sentence with inserted empty prepositions: "*..., ∅ we too are having to endure ∅ the greenhouse effects*".

("prp" in column "source-side" in table 1)

- its governing verb or noun (column "g.verb")
- the governed noun and its syntactic function in relation to its governor (col. "func,noun")

These source-side features, extracted from dependency parses (Choi and Palmer, 2012), are then projected to the target-side based on the word alignment (column "projected source-side"). Using source-side projections to identify the governor on the target-side eliminates the need to parse the disfluent MT output.

Finally, we use distributional subcategorization information as our third feature type (column "target-side subcat"). Relying on distributional subcategorization information (cf. section 6.1), we provide subcategorizational preferences for the observed verb in the form of *verb-preposition-case* tuples. The grammatical case indicates whether the noun is predominantly used as subject or direct/indirect object with an empty preposition. From the tuples, the system can learn, for example, that *unter etwas leiden* is a lot more plausible than *∅ etwas leiden*, even though the English sentence contains no preposition (*to endure sth.*). For each preposition, including ∅, we list how often the verb occurred with the respective preposition-case combination, with values ranging from 0 (no evidence) to 5 (high amount of observations); table 1 only shows three of these pairs.

From this training example, the model can learn that the second place-holder, even though aligned to an empty preposition governing an object on the English side, is not likely to be realized as a direct object as there is no evidence of the verb *leiden* (*to suffer*) with an accusative object, but a strong preference for the preposition *unter+Dat*. The projected noun (*Treibhauseffekt*) should rule out the possibility of ∅-Nom, as it is an unlikely subject of *leiden*. On the other hand, for the first place-holder preposition, all features point to a realization as ∅-Nom (subject). This example illustrates how the features can bridge the gap between the verb *leiden* and the place-holder to be realized as *unter* (middle part of the sentence omitted in the table).

In addition to tuples of the form *verb-preposition-case*, we also use *noun-noun*$_{genitive}$ tuples (not shown in table 1) to help the system decide whether two adjacent nouns headed with a place-holder should be realized as a *noun-noun*$_{genitive}$ construction (equivalent with English *noun-of-noun*), a *noun-prep-noun* construction or as two adjacent (subcategorized) NPs, for example NP$_{Acc}$ NP$_{Dat}$ (direct/indirect object).

5.2 Evaluation of Prediction Accuracy

The success of generating-prepositions in SMT depends to a large extent on the quality of the prediction component. Before beginning with the MT experiments, we thus evaluate the quality of predicting prepositions on clean data, the tuning-set.

We use the Wapiti toolkit (see section 6.1) to train a CRF to predict prepositions. We opted for a sequence model to take into account decisions from previous positions. Even though it only looks at previous decisions on bigram-level, the annotation of *case* on all elements of noun phrases should prevent that two adjacent noun phrases be assigned the same value for *case*.

Table 2 shows the performance of predicting prepositions on clean data. In the column "prep+case", we evaluate the accuracy of the prediction of both the preposition and its grammatical case, whereas the column "prep" gives the accuracy when only looking at the predicted preposition. We compare a model using source-side and projected source-side features (1) and a model with additional subcategorization information (2). Source-side information and its target-side pro-

	Features	prep+case	prep
1	basic + source	73.58	85.76
2	basic + source + subcat	73.42	85.78

Table 2: Results on clean data (3000 sentences).

prep	acc.	top-3 predicted (freq)
∅	95.17	∅ (10235), in (134), von (95)
in	79.19	in (1123), ∅ (170), von (21)
vor	77.14	vor (81), ∅ (10), bei (3)
nach	68.70	nach (90), ∅ (22), in (4)
zu	64.67	zu (238), ∅ (60), in (21)
an	61.09	an (179), ∅ (47), in (22)
unter	60.71	unter (34), ∅ (12), von (4)
auf	59.56	auf (215), ∅ (59), in (32)
aus	55.38	aus (72), ∅ (25), von (19)
wegen	22.22	wegen (4), für (4), ∅ (3)

Table 3: Individual prediction results.

jection are crucial – without source-information, content-conveying prepositions would need to be guessed – the addition of subcategorization information does not lead to further gains, though.

Table 3 lists the prediction results for some of the prepositions to be modeled, ranging from 95% to 22%. The realization as empty preposition constitutes by far the majority. In the list of the top-3 predicted prepositions, it becomes obvious that the realization as ∅ instead of an overt preposition is also the most frequent error; similarly, the prepositions *von/in* (*of/in*), all high-frequency prepositions, are often output instead of the correct preposition.

6 Experiments and Evaluation

Here, we present the setup and results of our experiments. In addition to the traditional metric BLEU, we assess the quality of the translated prepositions for a subset where relevant elements (verb, noun) match with the reference. Finally, we discuss some examples before concluding the paper.

6.1 Data and Experimental Setup

We trained a standard phrase-based Moses system on 4.3M lines of EN–DE data (WMT'14) with a 10.3M sentence language model. For the lemmatized representation of the morphology-aware SMT system, the German part was parsed with BitPar (Schmid, 2004) and analyzed with the morphological tool SMOR (Schmid et al., 2004). The models for predicting inflectional features and prepositions were built with the Wapiti toolkit (Lavergne et al., 2010). The inflectional models (*case, number, gender strong/weak*) were trained on lemma and tag information of the German part of the parallel data. The models to predict prepositions were trained on half of the parallel data due to the considerably larger amount of labels that can be predicted. The subcategorization tuples were extracted from German web data (Scheible et al. (2013), Faaß and Eckart (2013)) and Europarl. We used WMT'13 as tuning and WMT'14 as test sets[3].

6.2 Evaluation with BLEU

Table 4 shows the results of experiments with the baseline system (a), a morphology-aware SMT system with no special treatment for prepositions[4]. As a variant of the baseline system (b), we removed all prepositions from the translation output to be re-predicted. This does not lead to much change in BLEU, illustrating that the prediction step itself is not harmful. However, only changing existing prepositions is not sufficient and it is not possible to model empty vs. overt prepositions.

Table 5 shows results for the variants of the place-holder systems. Using a basic place-holder (□) representation (S1) leads to a considerably drop in relation to the baseline in table 4. Annotating the place-holder with *case* (S2) leads to an improvement of ca. 0.4, indicating that the abstract representation of the place-holders plays a significant role here.

In (S3), we mark whether the preposition is governed by a verb or a noun, to no avail. As an extension, we annotate the status of the place-holder: subcategorized or non-subcategorized in (S4), which seems to slightly help, even though the observed differences are very small. Assuming that functional prepositions contribute only little in terms of meaning, only subcategorized prepositions are represented by place-holders, whereas non-functional prepositions are kept. Again, we show two variants: in (S5a), all prepositions are re-predicted, while in (S5b), the forms of non-functional prepositions in the MT output are kept and only those for functional prepositions are predicted – this last result reaches the baseline level.

While none of the variants outperforms the baseline, we consider the results encouraging as they illustrate (i) that the representation of prepositions during the translation step considerably influences the MT quality (S2) and (ii) that applying the prediction step to a carefully selected subset of prepo-

[3]In the current version, we only work with the 1-best output of the MT system, and do not consider the n-best list.

[4]For comparison, Baseline$_{surface}$ shows the score for a non-morphology-aware system operating on surface forms.

System	Prepositions	BLEU	CRF
Baseline$_{surface}$	–	16.84	–
Baseline (a)	–	17.38	–
Baseline (b)	re-predict	17.36 17.31	src src+subcat

Table 4: Baseline variants (3003 sentences).

	Representation of place-holders	BLEU source	BLEU src+sub
S1	□	16.81	16.77
S2	□+Case	17.23	17.23
S3	□+Case+(V|N)	16.91	16.89
S4	□+Case+(V|N)+subcat	17.09	17.08
S5a	□+Case+(V|N): functional prp +Case+(V|N): non-func.	17.12	17.06
S5b	□+Case+(V|N): functional prp +Case+(V|N): non-func.	17.29	17.29

Table 5: Results for place-holder systems.

sitions improves the results (S5a vs. S5b).

6.3 Evaluation of Prepositions

BLEU is known to not capture subtle differences between two translation systems very well. Thus, we present a second evaluation in which we analyze the translation accuracy of prepositions.

It is difficult to automatically assess the quality of the translation of prepositions as the choice of a preposition depends on its context, mainly the verbs and/or nouns it occurs with. It is not sufficient to compare the prepositions occurring in the reference translation with those in the translation output, as the used verbs/nouns or even the entire structure of the sentence might differ. We will thus restrict the evaluation to cases where the relevant parts, namely the governing verb and the noun governed by the preposition are the same in the reference sentence and in the translation output[5]: in such cases, an automatic comparison of the preposition in the MT output with the preposition in the reference sentence is possible.

To obtain the set for which to evaluate the prepositions, we took each preposition in the reference sentence[6] governing a proper noun or named entity. The governing verb is identified relying on dependency parses of the reference translation. For extracting the equivalents of the relevant parts (preposition, noun, verb) in the translation output, we made use of the alignments with the English source sentence as pivot. The matching is made on lemma-level.

[5]We ignore PPs governed by nouns (such as *N von/an N* (*N of N*)) as they are often equivalent with genitive structures.

[6]The preposition needs to be in the group of the 17 prepositions which are subject of modeling in this work.

	BL	S2	S5
verb$_{MT}$ = verb$_{REF}$	502	469	503
verb$_{MT}$ = verb$_{REF}$, noun$_{MT}$ = noun$_{REF}$	270	260	271

Table 6: Subsets where governing verb/governed noun are the same in MT output and reference.

	BL	S2	S5a	S5b
verb$_{MT}$ = verb$_{REF}$	245 48.8%	233 49.7%	261 51.9%	250 49.7%
verb$_{MT}$ = verb$_{REF}$, noun$_{MT}$ = noun$_{REF}$	179 66.3%	174 66.9%	188 69.4%	178 65.7%

Table 7: Percentage of correct prepositions for the subsets from table 6.

Table 6 gives an overview of the amount of cases where the reference contains a preposition and its noun and governing verb are the same in the MT output; in the set of 3003 sentences, this is the case for a subset of 270 (baseline), 260 (S2, the best place-holder-only system) and 271 (S5). Note that the slightly less prep-noun-verb triples of S2 that match the reference compared to the baseline are not per-se a sign for inferior translation quality as we did not consider the possibility of synonymous translations.

Table 7 shows the amount of prepositions for the respective subsets that were considered correct, i.e. match with the reference. While the difference is very small, the percentage of correct prepositions is slightly higher for the systems S2/5a. Systems 5a/b are based on the same MT output; however, 5a fares better in this evaluation even though 5b had a higher BLEU score. We thus assume that BLEU did not improve based on the examined subset.

This analysis also shows that the translation quality of prepositions is a problem in need of more attention[7]. It has to be noted, though, that this evaluation only gives partial insights into the performance of the systems. The main problem is that the evaluation is centered around prepositions in the reference translation, which often is (structurally) different from the source sentence and consequently also the translation output. Thus, sentences with prepositions in the translation, but not in the reference, are not considered. Nevertheless, we regard this evaluation as suitable to evaluate the correctness of prepositions in an automatic way.

6.4 Examples

Here, we discuss outputs from the baseline and system 2 (cf. table 5) that cover the different syn-

[7]In some cases however, prepositions in the MT output are acceptable even if they do not match with the reference.

1	SRC	... malmon 's team will have to improve **on** recent performances .
	BL	... malmon das Team wird **über** die jüngsten Leistungen zu verbessern. *... malmon the team will* **over** *the recent performances improve.*
	NEW	... malmon das Team hat ∅ die jüngsten Leistungen zu verbessern . *... malmon the team has-to ∅ the recent performances improve*
	REF	... muss sich das Malmon-Team im Vergleich zu den vergangenen Auftritten auf jeden Fall steigern . *... must -refl- the malmon-team in comparison to the past performances in any case improve.*
2	SRC	outer space offers many possibilities for studying ∅ substances under extreme conditions ...
	BL	in den Weltraum bietet viele Möglichkeiten für das Studium ∅ Stoffe unter extremen Bedingungen ... *in the space offers many possibilities study$_{noun}$ ∅ substances under extreme conditions ...*
	NEW	der Raum bietet viele Möglichkeiten zum Studium **von** Stoffen unter extremen Bedingungen ... *in the space offers many possibilities for study$_{noun}$* **of** *substances under extreme conditions ...*
	REF	Das Weltall bietet viele Möglichkeiten, Materie unter extremen Bedingungen zu studieren ... *the universe offers many possibilities , substances under extreme conditions to study ...*
3	SRC	nowadays there are specialists **in** renovation to suit the needs of the elderly.
	BL	heutzutage gibt es Spezialisten **in** der Renovierung der Bedürfnisse der älteren Menschen. *nowadays there are specialists* **in** *the renovation of the needs of the elderly.*
	NEW	heutzutage gibt es Spezialisten **für** Renovierung , die die Bedürfnisse der älteren Menschen. *nowadays there are specialists* **for** *renovation, that the needs of the elderly.*
	REF	heute gibt es auch **für** den altersgerechten Umbau Spezialisten . *tody there are also for the age-appropriate renovation specialists.*
4	SRC	... what role the giant planet has played **in** the development of the solar system.
	BL	... welche Rolle der riesige Planet gespielt hat, **in** der Entwicklung des Sonnensystems. *... which role the giant planet played has,* **in** *the development of-the solar system.*
	NEW	... welche Rolle der riesige Planet gespielt hat **bei** der Entwicklung des Sonnensystems. *... which role the giant planet played has* **in** *the development of-the solar system.*
	REF	... welche Rolle der Riesenplanet bei der Entwicklung des Sonnensystems gespielt hat . *... which role the giant-planet in the development of the solar-system played has.*

Table 8: Example sentences.

tactic phenomena, namely different types of structural differences in source and target language, referred to in the introductory sections.

In (1), the preposition *on* should not be translated, as the verb *verbessern* (*to improve*) subcategorizes a direct object (*Leistungen/performances*). While there is a preposition (*über*) in the baseline, no preposition is produced by the new system, leading to a correct translation. As the reference does not match with the MT output, this sentence is not counted in the evaluation from the previous section or given credit from BLEU, even though it improved over the baseline.

In (2), the constellation is opposite: with no preposition in the English sentence, the baseline output is missing a preposition, marked with ∅. Here, the German structure is different as the verb *studying* is expressed by a noun (*Studium*). In this construction, the phrase containing *Stoffe* (*substances*) needs to be expressed as the PP *von Stoffen* (*of substances*). Alternatively, a *noun-noun$_{genitive}$* structure is possible – our system is able to produce both versions.

In (3), the literal translation of *in* in the baseline is not grammatical and the translation does not express the meaning of the source sentence. The new translation contains the appropriate preposition *für* and also correctly reproduces the source sentence.

Similarly, the preposition *bei* in (4) is a better choice than *in* in the baseline, even though the baseline sentence is understandable. This sentence pair is counted in the evaluation from the previous section, as the verb (*gespielt*) and noun (*Sonnensystem*) each match with the reference translation.

7 Conclusion and Future Work

We presented a novel system with an abstract representation for prepositions during translation and a post-processing component for generating target-side prepositions. In this setup, we effectively combine relevant source-side and target-side features. By making use of an abstract representation and then assigning all subcategorized elements to their respective functions to be inflected accordingly, our method can explicitly handle structural differences in source and target language. We thus believe that this is a sound strategy to handle the translation of prepositions.

While the systems fail to improve over the baseline, our experiments show that a meaningful representation of prepositions is crucial for translation quality. In particular, the annotation of *case* resulted in the best of all placeholder-only systems –

this information can be considered as a "light" semantic annotation. Consequently, a more semantically motivated annotation representing the semantic class of a preposition (e.g. temporal, local) might lead to a more meaningful representation and remains an interesting idea for future work. Alternatively, integrating the generation step of the prepositions into the decoding process, e.g. following (Tsvetkov et al., 2013), might be another promising strategy.

In our evaluation we discussed typical problems arising when translating prepositions. Furthermore, we addressed the problem of automatically evaluating the quality of prepositions in sentences that are often structured differently than the reference sentence by considering only the respective relevant elements. As the translation of prepositions remains a difficult problem in machine translation, an automatic method that takes into account both the morpho-syntactic as well as the semantic aspects of the realization of prepositions in their respective contexts is needed. In our evaluation, we take first steps into this direction.

Acknowledgments

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 644402, the DFG grants *Distributional Approaches to Semantic Relatedness* and *Models of Morphosyntax for Statistical Machine Translation* and a DFG Heisenberg Fellowship.

References

Agirre, Eneko, Aitziber Atutxa, Gorka Labaka, Mikel Lersundi, Aingeru Mayor, and Kepa Sarasola. 2009. Use of Rich Linguistic Information to Translate Prepositions and Grammatical Cases to Basque. In *Proceedings of EAMT*.

Choi, Jinho D. and Martha Palmer. 2012. Getting the Most out of Transition-Based Dependency Parsing. In *Proceedings of ACL*.

Eckle, Judith. 1999. *Linguistisches Wissen zur automatischen Lexikon-Akquisition aus deutschen Textcorpora*. Ph.D. thesis, Universität Stuttgart.

Faaß, Gertrud and Kerstin Eckart. 2013. SdeWaC – a Corpus of Parsable Sentences from the Web. In *Proceedings of GSCL*.

Fraser, Alexander, Marion Weller, Aoife Cahill, and Fabienne Cap. 2012. Modeling Inflection and Word-Formation in SMT. In *Proceedings of EACL*.

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

Lavergne, Thomas, Olivier Cappé, and François Yvon. 2010. Practical very large scale CRFs. In *Proceedings of ACL*.

Lo, Chi-kiu and Dekai Wu. 2011. MEANT: An inexpensive, high-accuracy, semi-automatic Metric for Evaluating Translation Utility via Semantic Frames. In *Proceedings of ACL*.

Naskar, Sudip Kumar and Sivaji Bandyopadhyay. 2006. Handling of Prepositions in English to Bengali Machine Translation. In *Proceedings of ACL-SIGSEM*.

Rozovskaya, Alla and Dan Roth. 2013. Joint Learning and Inference for Grammatical Error Correction. In *Proceedings of EMNLP*.

Scheible, Silke, Sabine Schulte im Walde, Marion Weller, and Max Kisselew. 2013. A Compact but Linguistically Detailed Database for German Verb Subcategorisation relying on Dependency Parses from a Web Corpus. In *Proceedings of WaC*.

Schmid, Helmut, Arne Fitschen, and Ulrich Heid. 2004. SMOR: a German Computational Morphology Covering Derivation, Composition, and Inflection. In *Proceedings LREC 2004*.

Schmid, Helmut. 2004. Efficient Parsing of Highly Ambiguous Context-Free Grammars with Bit Vectors. In *Proceedings of COLING*.

Shilon, Reshef, Hanna Fadida, and Shuly Wintner. 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*.

Toutanova, Kristina, Hisami Suzuki, and Achim Ruopp. 2008. Applying Morphology Generation Models to Machine Translation. In *Proceedings of ACL*.

Tsvetkov, Yulia, Chris Dyer, Lori Levin, and Archna Bhatia. 2013. Generating English Determiners in Phrase-Based Translation with Synthetic Translation Options. In *Proceedings of WMT*.

Weller, Marion, Alexander Fraser, and Sabine Schulte im Walde. 2013. Using Subcategorization Knowledge to Improve Case Prediction for Translation to German. In *Proceedings of ACL*.

Weller, Marion, Sabine Schulte im Walde, and Alexander Fraser. 2014. Using Noun Class Information to Model Selectional Preferences for Translating Prepositions in SMT. In *Proceedings of AMTA*.