Addressing Problems across Linguistic Levels in SMT:
Combining Approaches to Model Morphology, Syntax and Lexical Choice

Marion Weller-Di Marco[1,2], Alexander Fraser[2], and Sabine Schulte im Walde[1]

[1]Institut für Maschinelle Sprachverarbeitung, Universität Stuttgart
[2]Centrum für Informations- und Sprachverarbeitung, LMU München
{dimarco|schulte}@ims.uni-stuttgart.de
fraser@cis.lmu.de

Abstract

Many errors in phrase-based SMT can be attributed to problems on three linguistic levels: morphological complexity in the target language, structural differences and lexical choice. We explore combinations of linguistically motivated approaches to address these problems in English-to-German SMT and show that they are complementary to one another, but also that the popular verbal pre-ordering can cause problems on the morphological and lexical level. A discriminative classifier can overcome these problems, in particular when enriching standard lexical features with features geared towards verbal inflection.

1 Introduction and Motivation

Many of the errors occurring in SMT can be attributed to problems on three linguistic levels: morphological richness, structural differences between source and target language, and lexical choice. Often, these categories are intertwined: for example, the syntactic function of an argument can be expressed on the morphological level by grammatical case (e.g. in German), or on the syntactic level through word ordering (such as SVO in English).

This paper addresses problems across the three linguistic levels by combining established approaches which were previously studied only independently. We explore system variants that combine target-side morphological modeling, structural adaptation between source and target side and a discriminative lexicon enriched with features relevant for support verb constructions and verbal inflection. We show that the components targeting the different linguistic levels are complementary, but also that applying only verbal pre-ordering can introduce problems on the morpho-lexical level; our

experiments indicate that a discriminative classifier can overcome these problems.

In the following, we present some main strategies to address the linguistic levels individually.

Morphology Inflection is one of the main problems when translating into a morphologically rich language. It is subject to local restrictions such as agreement in nominal phrases, but also depends on sentence-level interactions, such as verb-subject agreement, or the realization of grammatical case.

Target-side morphology can be modeled through computation of inflectional features and generation of inflected forms (Toutanova et al., 2008; Fraser et al., 2012), by means of synthetic phrases to provide the full set of word inflections (Chahuneau et al., 2013), or by introducing agreement restrictions for consistent inflection (Williams and Koehn, 2011).

Syntax Different syntactic structures in source and target language are problematic as they are hard to capture by word alignment, and long-distance reorderings are typically also disfavoured in phrase-based SMT. Hierarchical systems can bridge gaps up to a certain length, possibly enhanced by explicit modeling, e.g. Braune et al. (2012).

An alternative method, especially for phrase-based systems, is source-side reordering: in a pre-processing step, the source-side data is arranged such that it corresponds to the target-side structure. This improves the alignment and does not require long-distance reordering during decoding, see e.g. Collins et al. (2005) and Gojun and Fraser (2012).

Lexicon Problems on the lexical level are diverse and include word sense disambiguation, selectional preferences and the translation of multi-word structures. Many approaches rely on rich source-side features to provide more context for decoding, e.g. Carpuat and Wu (2007), Jeong et al. (2010), Tamchyna et al. (2014), Tamchyna et al. (2016).
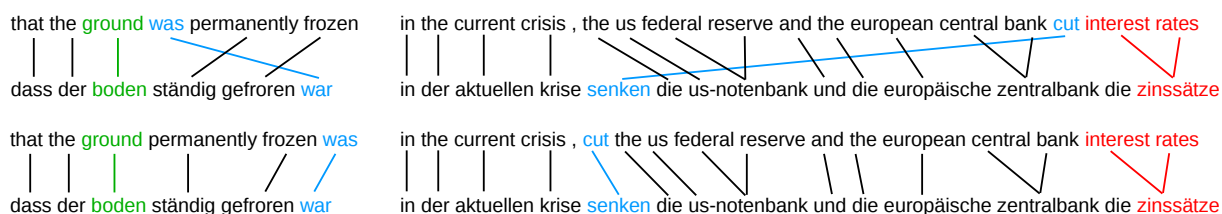
that the ground was permanently frozen

dass der boden ständig gefroren war

in the current crisis , the us federal reserve and the european central bank cut interest rates

in der aktuellen krise senken die us-notenbank und die europäische zentralbank die zinssätze

that the ground permanently frozen was

dass der boden ständig gefroren war

in the current crisis , cut the us federal reserve and the european central bank interest rates

in der aktuellen krise senken die us-notenbank und die europäische zentralbank die zinssätze

Figure 1: Verbal reordering in the training data: *verb-final position* (left) and *verb-second position* (right).

Combining Approaches Individual strategies aiming at one linguistic level are established and usually improve translation, but it is not clear (i) whether individual gains add up when combining approaches and (ii) how individually targeting one linguistic level impacts other levels. We address these questions for the combined strategies of *source-side reordering* (pre-processing), *discriminative classifier* (at decoding time) and *target-side generation* of nominal inflection (post-processing). For (ii), we focus on source-side reordering and investigate whether introducing German clause ordering in the English data entails new problems: while in "regular" English verbs and their arguments are close to each other, they can be separated by large distances in the German-structured English.

Reordering improves translation quality, but separating the verb from its arguments has also negative consequences. First, the agreement in number between verbs and subjects is impaired because subjects and verbs are separated (Ramm and Fraser, 2016). Second, there can be a negative effect on the lexical level, for example when translating multi-word expressions. Consider the phrase *to cut interest rates*: if the parts occur close to each other, there is enough context to translate *cut* into *senken* ('to decrease'). However, with too large a gap between *cut* and *interest rates*, it becomes difficult to disambiguate *cut*, leading to the wrong translation *schneiden* ('to cut with a knife').

2 Morpho-Syntactic Modeling

This section outlines the pre- and post-processing steps for morpho-syntactic modeling.

Morphology Nominal morphology is handled by an inflection prediction process which first translates into an underspecified stemmed representation and then generates inflected forms in a post-processing step (Fraser et al., 2012). The stemmed representation is enriched with translation-relevant features, such as number on nouns, to ensure that number as expressed on the source side is preserved during translation. To re-inflect the stemmed SMT output, inflectional features are predicted with classifiers using the values in the stem-markup as input. The inflected forms are then generated from the stem+feature pairs using a morphological resource.

Reordering English verbs are moved to the expected German position, following the rules in Gojun and Fraser (2012). The resulting structure is fundamentally different from "regular" English, as illustrated in figure 1. The left side shows the movement of an English verb to the *verb-final* position in a subordinated clause, inserting a gap between verb and subject. This might well have a negative impact on subject-verb agreement: while *was* is obviously singular, modal verbs and verbs in past tense require context to determine number. The right side depicts *verb-second position*, where the finite verb is moved to the second constituent.

Long-distance reorderings as in this example are not uncommon and their benefit on verbal translation is intuitively clear. However, reordering comes at the price of separating the verb and its direct object. This is particularly problematic when verb and object form a multi-word expression: (parts of) the expression cannot be translated literally, but need to take into account the context. When the source-side is reordered, the system has better word alignments of verbal translations, but less context to distinguish between translation senses. Furthermore, non-finite verbs in compound tenses (*have/would ... cut*) go to the end of the clause, separating auxiliaries and full verbs. As German auxiliaries for past tense depend on the verb, a separation can impair the selection of the auxiliary.

3 Context Features for Lexical Modeling

Rich source-side context features provide information on the lexical level, but also for morpho-syntactic concerns such as number agreement or auxiliary choice. We use a discriminative classifier (VowpalWabbit[1]), which is integrated into

---

[1] https://github.com/JohnLangford/vowpal_wabbit/wiki

626

word	pos	lemma	associated verb/noun	rel-ation	svc
cut	vvd	cut	rate	dobj	250
the	dt	the	–	–	–
us	np	us	reserve	nn-mod	–
federal	np	federal	reserve	nn-mod	–
reserve	np	reserve	cut	nsubj	–
...					
interest	nn	interest	rate	nn-mod	–
rates	nns	rate	cut	dobj	–

Table 1: Subject/object relations and support verb status on the reordered sentence from figure 1.

.

the Moses framework, in order to score translation rules using rich source context information outside of the applied phrase (Tamchyna et al., 2014). We employ different feature types for source context:

Standard Features on the source-side comprise part-of-speech tags and lemmas within the phrase and a context window (5 for tags, 3 for word/lemma). Information across larger gaps is captured by dependency relations such as verb-object pairs or verb-subject pairs, cf. columns 4 and 5 in table 1. On the target-side, lemmas and part-of-speech tags for the current phrase are given.

Support Verb Constructions are formed by a verb and a predicative noun, e.g. *make a contribution*. Typically, the verb does not contribute its full meaning, and thus cannot be translated literally. Cap et al. (2015) improved German-English phrase-based SMT by annotating support verb status on source-side verbs, which essentially divides verbs into two groups: "non-literal use" in a support verb construction, and "literal use" otherwise. The set of support verb constructions consists of highly associated noun+verb tuples. Cap et al. (2015) opted for a hard annotation by adding markup. Instead, we add a classifier feature and compare two variants:

(i) setting the feature to a *binary support verb status* (yes/no) for a fixed set of tuples (using a log-likelihood threshold of 1000, as in Cap et al. (2015)). There is no dependency information in this variant, only the basic features lemma and POS-tag.

(ii) annotating the *degree of relatedness* between verb and noun (i.e. log-likelihood score) in addition to the dependency information, see rightmost column in table 1. Verb-noun tuples are grouped into sets based on their degree of association (e.g. log-likelihood score between 250 and 500). This allows us to always annotate support verb status, instead of arbitrarily deciding on a threshold.

Number and Tense Information The complexity of verbal inflection is generally difficult to capture, in particular when complex interactions between several verbs are involved. Lóaiciga et al. (2014) investigate rich source-side features in factored MT and improved the translation of tense for English–French MT. Reordering might make verbal inflection even more difficult, with regard to subject-verb agreement and the choice of auxiliaries. While the number of verbs in present tense is often obvious (*goes* vs. *go*), verbs in past tense (*went*) or progressive form (*going*) require the subject for disambiguation. Number, as derived from the subject, is used as an extra feature for verbs.

As the reordering complicates the processing of a compound past (e.g. *has ... gone, did ... buy*), we annotate the status of *past* vs. *non-past*, as well as the associated other verb. This aims at providing information to decide for the correct tense and to select the correct auxiliary (*sein*: 'to be' vs. *haben*: 'to have') for German present/past perfect.

4 Experiments and Results

This section presents the results of combining the strategies for the three linguistic levels.

Data and Resources All systems are built using the Moses phrase-based framework. The translation model is based on 4.592.139 parallel sentences; and 45M sentences (News14+parallel data) are used to train a 5-gram language model. We use NewsTest'13 (3000 sentences) and News Test'14 (3003 sentences) for tuning and testing. The linguistic processing for inflection prediction includes parsing (Schmid, 2004) and morphological analysis/generation (Schmid et al., 2004). To predict the features for nominal inflection, CRF sequence models (Lavergne et al., 2010) are trained on the target-side of the parallel data. The reordering rules from Gojun and Fraser (2012) are applied to parsed English data (Charniak and Johnson, 2005).

We use a version of Moses with the integrated discriminative classifier VowpalWabbit (Tamchyna et al., 2014)[2]. Training examples are extracted from the parallel data based on phrase-table entries. In order to keep the amount of training examples manageable, the phrase-table is reduced with sigtest-filtering with the setting *-l a+e -n 30*.[3] We run 50 training iterations and apply early-stopping on the development set to identify the optimal model.

---

[2]github.com/moses-smt/mosesdecoder/tree/master/vw
[3]All experiments use sigtest-filtered phrase-tables.

system	basic	VW-1 pos/lem	VW-2 pos/lem/dep
Surface	19.45	19.81*	19.90*
Surface V-Reordered	19.71*	20.24*	20.27*
MorphSys	19.81*	19.80*	19.93*
MorphSys V-Reordered	20.08*	20.51*	20.50*

Table 2: Morpho-syntactic and lexical strategies. *: significantly better than Surface-basic (19.45)

system	VW-2	VW-1 +threshold	VW-2 +degree
MorphSys	19.93	20.07	19.98
MorphSys V-Reordered	20.50	20.40	20.46

Table 3: Annotating support verb status.

Morpho-Syntactic and Lexical Strategies The column "basic" in table 2 shows the results for combining strategies at the morpho-syntactic level: "Surface" refers to a baseline system trained on surface forms; "MorphSys" denotes the inflection prediction system; "V-Reordered" refers to systems built on reordered source-side data. Combining the two strategies adds up to a statistically significant gain of 0.63 between the basic system (19.45) and the system with morphological modeling and source-side reordering (20.08).

The columns show the effect of the discriminative model. Classifier VW-1 uses *word/ lemma/pos* information; VW-2 is extended with dependency relations. The difference between the two classifiers is small. Compared to the basic surface system, the "MorphSys" system does not gain much; presumably because the classifier contributes to the morphological level for the surface system, such as triggering consistent inflection, which is already an integral part in the "MorphSys" system. Systems built on reordered source-side data tend to benefit more from the additional lexical information, which confirms our hypothesis that verbal reordering is problematic at the lexical level. Combining all strategies leads to the overall best result.

Support Verb Constructions and Verb Features The two systems with inflection prediction are enriched with information about support verb constructions, in form of a binary annotation to the features of VW-1, or by annotating the degree of association to the features of VW-2, cf. table 3. Both variants do not improve over the systems with classifiers VW-1 or VW-2. Since support verb constructions are already indirectly contained in the

system	VW-2	VW-1 +num	VW-2 +num	VW-2 +num +tense
MorphSys	19.93	20.00	20.00	20.02
MorphSys V-Reordered	20.50	20.60	20.57	20.62

Table 4: Annotating number and tense information.

	better	worse	equal
number agreement	20	2	4
auxiliary (past/passive)	11	5	2
tense	4	4	2
missing/extra verb	61	20	14
none of the above	0	0	17

Table 5: Manual evaluation of 155 sentences.

dependency information, the explicit annotation does not seem to provide extra knowledge.

The reordered and non-reordered "MorphSys" systems are extended with verbal features, leading to minor improvements over classifier VW-2[4], cf. table 4. To examine the effect of modeling tense and number, we compared the output of system VW-2 (reordered) with the enriched system (reordered VW-2 +Num+Tense). As test set, we extracted sentences containing at least one difference in verb translations, and additionally restricted the source sentence length to 8-20 words. After removing sentences with only lexically different verbs, 155 sentences remained. 3 native speakers of German manually rated each pair of differently translated verbs (ignoring all other words) with respect to the following categories:

- **Number agreement:** subject and verb agree in *number*. The value "equal" can apply if the subject is translated differently, e.g. *research shows* vs. *studies show*.

- **Auxiliary:** presence, absence and choice of auxiliary, e.g. *sein* ('to be') vs. *haben* ('to have') as auxiliary for past tense.

- **Tense:** the translation reproduces the tense in the source-sentence, as well as the technical correctness for compound tenses, e.g. *has done* vs. *has did* vs. ∅ *done*.

- **Missing/extra verb:** refers to the number of full verbs in the sentence. In this category, it is mostly the case that verbs are missing, but it also happens that superfluous verbs appear in a translation.

---

[4]Even though small, the difference between 20.50 and 20.62 is statistically significant with pair-wise bootstrap resampling with sample size 1,000 and a p-value of 0.05.

SRC	i really feel that he should follow in the footsteps of the other guys .
reordered	i really feel that <u>he</u> in the footsteps of the other guys follow <u>should</u> .
VW2	ich bin wirklich der Meinung , dass <u>er</u> in die Fußstapfen der anderen Jungs folgen **sollten**$_{PL}$. *i am really of-the opinion , that <u>he</u> in the footsteps of the other guys follow **should***
+NumTense	ich bin wirklich der Meinung , dass <u>er</u> in die Fußstapfen der anderen Jungs folgen **sollte**$_{SG}$. *i am really of-the opinion , that <u>he</u> in the footsteps of the other guys follow **should** .*

Table 6: Example for improvement of number agreement due to the number annotation on the verb *should*.

SRC	television footage revealed how numerous ambulances and police cars arrived at a terminal .
reordered	television footage revealed how numerous ambulances and police cars at a terminal <u>arrived</u> .
VW2	das Fernsehen zeigte Bilder , wie zahlreiche Rettungswagen und Polizei Autos an einem Terminal . *the television showed images , how numerous ambulances and police cars at a terminal .*
+NumTense	das Fernsehen zeigte Bilder , wie zahlreiche Rettungswagen und Polizei Autos an einem Terminal **angekommen** . *the television showed images , how numerous ambulances and police cars at a terminal **arrived** .*

Table 7: Example for the addition of a missing verb.

SRC	it would thus be suitable to assist illegal immigration into the usa .
reordered	it <u>would</u> thus suitable <u>be</u> illegal immigration into the usa to assist .
VW2	es **wäre** daher geeignet **sein** , die illegale Einwanderung in die USA zu unterstützen . *it **would-be** thus suitable **be** , the illegal immigration into the usa to assist .*
+NumTense	es **wäre** daher ideal , illegale Einwanderung in die USA zu unterstützen . *it **would-be** thus ideal , illegal immigration into the usa to assist .*

Table 8: Example for the removal of a superfluous verb.

- **None of the above:** refers mostly to translation of poor quality, so that verb translations cannot be analyzed properly.

The results in table 5 show that the enriched system is better with regard to verb-subject number agreement, choice of auxiliary and the number of missing/superfluous verbs.

The annotation of number is very straightforward, as it is a single piece of information which is easy to obtain: its effect is illustrated in table 6, where the enriched system produces the correctly inflected form *sollte*, whereas the other system has no access to the subject's number at the end of the sentence and incorrectly outputs a plural form. The modeling of tense features is more complex, because several verbs may be involved, and their effect cannot be explained as easily as in the number example. We assume that the richer annotation results in slightly more precise estimations that promote better translations. For example, the output produced by the enriched system in table 7 contains a verb that is missing in the other system. Even though it is not technically well-formed (past participle without auxiliary), this constitutes an improvement. On the other hand, the VW-2 system in table 8 produces the extra verb *sein* ('be'), at the position corresponding to the source-side *be*. However, the verb *wäre* already is a finite verb with the meaning *would be*, making the second verb redundant. In the enriched version, *be* is annotated with its related verb *would*, and thus might trigger a preference for a translation without verb in this context, as *would→wäre* is already sufficient.

5 Conclusion

We presented and combined established approaches to address the linguistic levels *Morphology*, *Syntax* and *Lexical Choice* in phrase-based SMT. By comparing combinations of strategies to address these problems for English-to-German SMT, we showed that they are complementary to one another. We pointed out that verbal reordering can introduce problems on the morphological and lexical level. Our results indicate that it is possible to overcome these problems by using a discriminative lexicon; enriching standard features with information for verbal inflection leads to a further improvement.

Acknowledgments

References

Fabienne Braune, Anita Gojun, and Alexander Fraser. 2012. Long Distance Reordering During Search for Hierarchical Phrase-based SMT. In *Proceedings of the 16th Annual Conference of the European Association for Machine Translation (EAMT)*, Trento, Italy.

Fabienne Cap, Manju Nirmal, Marion Weller, and Sabine Schulte im Walde. 2015. How to Account for Idiomatic German Support Verb Constructions in Statistical Machine Translation. In *Proceedings of the 11th Workshop on Multiword Expressions at NAACL.*, Denver, Colorado.

Marine Carpuat and Dekai Wu. 2007. Improving Statistical Machine Translation using Word Sense Disambiguation . In *Proceedings of the Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL 2007)*, Prague, Czech Republic.

Victor Chahuneau, Eva Schlinger, Noah A. Smith, and Chris Dyer. 2013. Translating into Morphologically Rich Languages with Synthetic Phrases. In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, Seattle, Washington.

Eugene Charniak and Mark Johnson. 2005. Coarse-to-fine n-best parsing and MaxEnt discriminative reranking. In *In Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics (ACL)*, Ann Arbour, Michigan.

Michael Collins, Philipp Koehn, and Ivona Kucerova. 2005. Clause Restructuring for Statistical Machine Translation. In *Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics (ACL)*, Ann Arbour, Michigan.

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

Anita Gojun and Alexander Fraser. 2012. Determining the Placement of German Verbs in English-to-German SMT. In *Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics (EACL)*, Avignon, France.

Minwoo Jeong, Kristina Toutanova, Chris Quirk, and Hisami Suzuki. 2010. A Discriminative Lexicon Model for Complex Morphology . In *Proceedings of the Ninth Conference of the Association for Machine Translation in the America (ACL)*, Uppsala, Sweden.

Thomas Lavergne, Olivier Cappé, and François Yvon. 2010. Practical very large scale CRFs. In *Proceedings the 48th Annual Meeting of the Association for Computational Linguistics (ACL)*, Uppsala, Sweden.

Sharid Lóaiciga, Thomas Meyer, and Andrei Popescu-Belis. 2014. English-French Verb Phrase Alignment in Europarl for Tense Translation Modeling. In *Proceedings of LREC 2014*, Reykjavik, Iceland.

Anita Ramm and Alexander Fraser. 2016. Modeling Verbal Inflection for English to German SMT. In *Proceedings of the First Conference of Machine Translation (WMT16)*, Berlin, Germany.

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

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

Aleš Tamchyna, Fabienne Braune, Alexander Fraser, Marine Carpuat, Hal Daume III, and Chris Quirk. 2014. Integrating a Discriminative Classifier into Phrase-based and Hierarchical Decoding. In *The Prague Bulletin of Mathematical Linguistics, No. 101, pages 29-41*.

Aleš Tamchyna, Alexander Fraser, Ondřej Bojar, and Marcin Junczys-Dowmunt. 2016. Target-Side Context for Discriminative Models in Statistical Machine Translation. In *Proceedings of ACL 2016*, Berlin, Germany.

Kristina Toutanova, Hisami Suzuki, and Achim Ruopp. 2008. Applying Morphology Generation Models to Machine Translation. In *Proceedings of ACL08-HLT*, Columbus, Ohio.

Philip Williams and Philipp Koehn. 2011. Agreement Constraints for Statistical Machine Translation into German . In *Proceedings of the 6th Workshop on Statistical Machine Translation (WMT)*, Edinburgh, UK.