Munich-Edinburgh-Stuttgart Submissions at WMT13: Morphological and Syntactic Processing for SMT

Marion Weller[1], **Max Kisselew**[1], **Svetlana Smekalova**[1], **Alexander Fraser**[2], **Helmut Schmid**[2], **Nadir Durrani**[3], **Hassan Sajjad**[4], **Richárd Farkas**[5]

[1]University of Stuttgart – (wellermn|kisselmx|smekalsa)@ims.uni-stuttgart.de

[2]Ludwig-Maximilian University of Munich – (schmid|fraser)@cis.uni-muenchen.de

[3]University of Edinburgh – dnadir@inf.ed.ac.uk

[4]Qatar Computing Research Institute – hsajjad@qf.org.qa

[5]University of Szeged – rfarkas@inf.u-szeged.hu

Abstract

We present 5 systems of the Munich-Edinburgh-Stuttgart[1] joint submissions to the 2013 SMT Shared Task: FR-EN, EN-FR, RU-EN, DE-EN and EN-DE. The first three systems employ inflectional generalization, while the latter two employ parser-based reordering, and DE-EN performs compound splitting. For our experiments, we use standard phrase-based Moses systems and operation sequence models (OSM).

1 Introduction

Morphologically complex languages often lead to data sparsity problems in statistical machine translation. For translation pairs with morphologically rich source languages and English as target language, we focus on simplifying the input language in order to reduce the complexity of the translation model. The pre-processing of the source-language is language-specific, requiring morphological analysis (FR, RU) as well as sentence reordering (DE) and dealing with compounds (DE). Due to time constraints we did not deal with inflection for DE-EN and EN-DE.

The morphological simplification process consists in lemmatizing inflected word forms and dealing with word formation (splitting portmanteau prepositions or compounds). This needs to take into account translation-relevant features (e.g. *number*) which vary across the different language pairs: while French only has the features *number* and *gender*, a wider array of features needs to be considered when modelling Russian (cf. table 6). In addition to morphological reduction, we also apply transliteration models learned from automatically mined transliterations to handle out-of-vocabulary words (OOVs) when translating from Russian.

Replacing inflected word forms with simpler variants (lemmas or the components of split compounds) aims not only at reducing the general complexity of the translation model, but also at decreasing the amount of out-of-vocabulary words in the input data. This is particularly the case with German compounds, which are very productive and thus often lack coverage in the parallel training data, whereas the individual components can be translated. Similarly, inflected word forms (e.g. adjectives) benefit from the reduction to lemmas if the full inflection paradigm does not occur in the parallel training data.

For EN-FR, a translation pair with a morphologically complex target language, we describe a two-step translation system built on non-inflected word stems with a post-processing component for predicting morphological features and the generation of inflected forms. In addition to the advantage of a more general translation model, this method also allows the generation of inflected word forms which do not occur in the training data.

2 Experimental setup

The translation experiments in this paper are carried out with either a standard phrase-based Moses system (DE-EN, EN-DE, EN-FR and FR-EN) or with an operation sequence model (RU-EN, DE-EN), cf. Durrani et al. (2013b) for more details. An operation sequence model (OSM) is a state-of-the-art SMT-system that learns translation and reordering patterns by representing a sentence pair and its word alignment as a unique sequence of operations (see e.g. Durrani et al. (2011), Durrani et al. (2013a) for more details). For the Moses systems we used the old train-model perl scripts rather than the EMS, so we did not perform Good-Turing smoothing; parameter tuning was carried out with *batch-mira* (Cherry and Foster, 2012).

[1]The language pairs DE-EN and RU-EN were developed in collaboration with the Qatar Computing Research Institute and the University of Szeged.

Proceedings of the Eighth Workshop on Statistical Machine Translation, pages 232–239,
Sofia, Bulgaria, August 8-9, 2013

1	Removal of empty lines
2	Conversion of HTML special characters like *"* to the corresponding characters
3	Unification of words that were written both with an *œ* or with an *oe* to only one spelling
4	Punctuation normalization and tokenization
5	Putting together clitics and apostrophes like *l '* or *d '* to *l'* and *d'*

Table 1: Text normalization for FR-EN.

Definite determiners Indefinite determiners	*la / l' / les* → *le* *un / une* → *un*
Adjectives	Infl. form → lemma
Portmanteaus	e. g. *au* → *à le*
Verb participles inflected for gender and number ending in *ée/és/ées*	Reduced to non-inflected verb participle form ending in *é*
Clitics and apostrophized words are converted to their lemmas	*d'* → *de*, *qu'* → *que*, *n'* → *ne*, ...

Table 2: Rules for morphological simplification.

The development data consists of the concatenated news-data sets from the years 2008-2011. Unless otherwise stated, we use all constrained data (parallel and monolingual). For the target-side language models, we follow the approach of Schwenk and Koehn (2008) and train a separate language model for each corpus and then interpolate them using weights optimized on development data.

3 French to English

French has a much richer morphology than English; for example, adjectives in French are inflected with respect to gender and number whereas adjectives in English are not inflected at all. This causes data sparsity in coverage of French inflected forms. We try to overcome this problem by simplifying French inflected forms in a pre-processing step in order to adapt the French input better to the English output.

Processing of the training and test data The pre-processing of the French input consists of two steps: (1) normalizing not well-formed data (cf. table 1) and (2) morphological simplification.

In the second step, the normalized training data is annotated with Part-of-Speech tags (PoS-tags) and word lemmas using RFTagger (Schmid and Laws, 2008) which was trained on the French treebank (Abeillé et al., 2003). French forms are then simplified according to the rules given in table 2.

Data and experiments We trained a French to English Moses system on the preprocessed and simplified constrained parallel data.

Due to tractability problems with word alignment, the 10^9 French-English corpus and the UN corpus were filtered to a more manageable size. The filtering criteria are sentence length (between 15 and 25 words), as well as strings indicating that a sentence is neither French nor English, or otherwise not well-formed, aiming to obtain a subset of good-quality sentences. In total, we use 9M parallel sentences. For the English language model we use large training data with 287.3M true-cased sentences (including the LDC Giga-word data).

System	**BLEU** (cs)	**BLEU** (ci)
Baseline	29.90	31.02
Simplified French*	29.70	30.83

Table 3: Results of the French to English system (WMT-2012). The marked system (*) corresponds to the system submitted for manual evaluation. (cs: case-sensitive, ci: case-insensitive)

We compare two systems: a baseline with regular French text, and a system with the described morphological simplifications. Results for the WMT-2012 test set are shown in table 3. Even though the baseline is better than the simplified system in terms of BLEU, we assume that the translation model of the simplified system benefits from the overall generalization – thus, human annotators might prefer the output of the simplified system.

For the WMT-2013 set, we obtain BLEU scores of 29,97 (cs) and 31,05 (ci) with the system built on simplified French (*mes-simplifiedfrench*).

4 English to French

Translating into a morphologically rich language faces two problems: that of asymmetry of morphological information contained in the source and target language and that of data sparsity.

In this section we describe a two-step system designed to overcome these types of problems: first, the French data is reduced to non-inflected forms (stems) with translation-relevant morphological features, which is used to built the translation model. The second step consists of predicting all necessary morphological features for the translation output, which are then used to generate fully inflected forms. This two-step setup decreases the complexity of the translation task by removing language-specific features from the translation model. Furthermore, generating inflected forms based on word stems and morphological features allows to generate forms which do not occur in the parallel training data – this is not possible in a standard SMT setup.

The idea of separating the translation into two steps to deal with complex morphology was introduced by Toutanova et al. (2008). Fraser et al. (2012) applied this method to the language pair English-German with an additional special focus on word formation issues such as the splitting and merging of portmanteau prepositions and compounds. The presented inflection prediction systems focuses on nominal inflection; verbal inflection is not addressed.

Morphological analysis and resources The morphological analysis of the French training data is obtained using RFTagger, which is designed for annotating fine-grained morphological tags (Schmid and Laws, 2008). For generating inflected forms based on stems and morphological features, we use an extended version of the finite-state morphology FRMOR (Zhou, 2007). Additionally, we use a manually compiled list of abbreviations and named entities (names of countries) and their respective grammatical gender.

Stemming For building the SMT system, the French data (parallel and monolingual) is transformed into a stemmed representation. Nouns, i.e. the heads of NPs or PPs, are marked with inflection-relevant features: *gender* is considered as part of the stem, whereas *number* is determined by the source-side input: for example, we expect source-language words in plural to be translated by translated by stems with plural markup. This stem-markup is necessary in order to guarantee that the number information is not lost during translation. For a better generalization, portmanteaus are split into separate parts: *au* $\rightarrow$ *à+le* (meaning, "to the").

Predicting morphological features For predicting the morphological features of the SMT output (*number* and *gender*), we use a linear chain CRF (Lavergne et al., 2010) trained on data annotated with these features using n-grams of stems and part-of-speech tags within a window of 4 positions to each side of the current word. Through the CRF, the values specified in the stem-markup (*number* and *gender* on nouns) are propagated over the rest of the linguistic phrase, as shown in column 2 of table 4. Based on the stems and the morphological features, inflected forms can be generated using FRMOR (column 3).

Post-processing As the French data has been normalized, a post-processing step is needed in order to generate correct French surface forms: split portmanteaus are merged into their regular forms based on a simple rule set. Furthermore, apostrophes are reintroduced for words like *le, la, ne, ...* if they are followed by a vowel. Column 4 in table 4 shows post-processing including portmanteau formation. Since we work on lowercased data, an additional recasing step is required.

Experiments and evaluation We use the same set of reduced parallel data as the FR-EN system; the language model is built on 32M French sentences. Results for the WMT-2012 test set are given in table 5. Variant 1 shows the results for a small system trained only on a part of the training data (Europarl+News Commentary), whereas variant 2 corresponds to the submitted system. A small-scale analysis indicated that the inflection prediction system tends to have problems with subject-verb agreement. We trained a factored system using additional PoS-tags with number information which lead to a small improvement on both variants.

While the small model is significantly better than the baseline[2] as it benefits more from the generalization, the result for the full system is worse than the baseline[3]. Here, given the large amount of data, the generalization effect has less influence. However, we assume that the more general model from the inflection prediction system produces better translations than a regular model containing a large amount of irrelevant inflectional information, particularly when considering that it can produce well-formed inflected sequences that are inaccessible to the baseline. Even though this is not reflected in terms of BLEU, humans might prefer the inflection prediction system.

For the WMT-2013 set, we obtain BLEU scores of 29.6 (ci) and 28.30 (cs) with the inflection prediction system *mes-inflection* (marked in table 5).

5 Russian-English

The preparation of the Russian data includes the following stages: (1) tokenization and tagging and (2) morphological reduction.

Tagging and tagging errors For tagging, we use a version of RFTagger (Schmid and Laws, 2008)

[2]Pairwise bootstrap resampling with 1000 samples.

[3]However, the large inflection-prediction system has a slightly better NIST score than the baseline (7.63 vs. 7.61).

SMT-output with stem-markup in bold print	**predicted features**	**generated forms**	**after post-processing**	**gloss**
avertissement<**Masc**><**Pl**>[N]	Masc.Pl	avertissements	avertissements	*warnings*
sinistre[ADJ]	Masc.Pl	sinistres	sinistres	*dire*
de[P]	–	de	du	*from*
le[ART]	Masc.Sg	le		*the*
pentagone<**Masc**><**Sg**>[N]	Masc.Sg	pentagone	pentagone	*pentagon*
sur[P]	–	sur	sur	*over*
de[P]	–	de	d'	*of*
éventuel[ADJ]	Fem.Pl	éventuelles	éventuelles	*potential*
réduction<**Fem**><**Pl**>[N]	Fem.Pl	réductions	réductions	*reductions*
de[P]	–	de	du	*of*
le[ART]	Masc.Sg	le		*the*
budget<**Masc**><**Sg**>[N]	Masc.Sg	budget	budget	*budget*
de[P]	–	de	de	*of*
le[ART]	Fem.Sg	la	la	*the*
défense<**Fem**><**Sg**>[N]	Fem.Sg	défense	défense	*défense*

Table 4: Processing steps for the input sentence *dire warnings from pentagon over potential defence cuts*.

that has been developed based on data tagged with TreeTagger (Schmid, 1994) using a model from Sharoff et al. (2008). The data processed by TreeTagger contained errors such as wrong definition of PoS for adverbs, wrong selection of gender for adjectives in plural and missing features for pronouns and adverbs. In order to train RFTagger, the output of TreeTagger was corrected with a set of empirical rules. In particular, the morphological features of nominal phrases were made consistent to train RFTagger: in contrast to TreeTagger, where morphological features are regarded as part of the PoS-tag, RFTagger allows for a separate handling of morphological features and POS tags.

Despite a generally good tagging quality, some errors seem to be unavoidable due to the ambiguity of certain grammatical forms in Russian. A good example of this are neuter nouns that have the same form in all cases, or feminine nouns, which have identical forms in singular genitive and plural nominative (Sharoff et al., 2008). Since Russian has no binding word order, and the case of nouns cannot be determined on that basis, such errors cannot be corrected with empirical rules implemented as post-processing. Similar errors occur when specifying the case of adjectives, since the suffixes of adjectives are even less varied as compared to the nouns. In our application, we hope that this type of error does not affect the result due to the following suppression of a number of morphological attributes including the case of adjectives.

	System	**BLEU** (ci)	**BLEU** (cs)
1	Baseline	24.91	23.40
	InflPred	25.31	23.81
	InflPred-factored	25.53	24.04
2	Baseline	29.32	27.65
	InflPred*	29.07	27.40
	InflPred-factored	29.17	27.46

Table 5: Results for French inflection prediction on the WMT-2012 test set. The marked system (*) corresponds to the system submitted for manual evaluation.

Morphological reduction In comparison to Slavic languages, English is morphologically poor. For example, English has no morphological attributes for nouns and adjectives to express gender or case; verbs have no gender either. In contrast, Russian is morphologically very rich – there are e.g. 6 cases and 3 grammatical genders, which manifest themselves in different suffixes for nouns, pronouns, adjectives and some verb forms. When translating from Russian into English, many of these attributes are (hopefully) redundant and are therefore deleted from the training data. The morphological reduction in our system was applied to nouns, pronouns, verbs, adjectives, prepositions and conjunctions. The rest of the POS (adverbs, particles, interjections and abbreviations) have no morphological attributes. The list of the original and the reduced attributes is given in Table 6.

Transliteration mining to handle OOVs The machine translation system fails to translate out-of-vocabulary words (OOVs) as they are unknown to the training data. Most of the OOVs are named entities and transliterating them to the target language script could solve this problem. The transliteration system requires a list of transliteration pairs for training. As we do not have such a list, we use the unsupervised transliteration mining system of Sajjad et al. (2012) that takes a list of word pairs for

Part of Speech	Attributes RFTagger	Reduced attributes
Noun	Type Gender Number Case $nom,gen,dat,acc,instr,prep$ Animate Case 2	Type Gender Number Case $gen,notgen$
Pronoun	Person Gender Number Case $nom,gen,dat,acc,instr,prep$ Syntactic type Animated	Person Gender Number Case $nom,notnom$
Verb	Type VForm Tense Person Number Gender Voice Definiteness Aspect Case	Type VForm Tense Person Number Voice Aspect
Adjective	Type Degree Gender Number Case Definiteness	Type Degree
Preposition	Type Formation Case	
Conjunction	Type Formation	Type Formation

Table 6: Rules for simplifying the morphological complexity for RU.

training and extracts transliteration pairs that can be used for the training of the transliteration system. The procedure of mining transliteration pairs and transliterating OOVs is described as follows: We word-align the parallel corpus using GIZA++ and symmetrize the alignments using the *grow-diag-final-and* heuristic. We extract all word pairs which occur as 1-to-1 alignments (Sajjad et al., 2011) and later refer to them as a *list of word pairs*. We train the unsupervised transliteration mining system on the list of word pairs and extract transliteration pairs. We use these mined pairs to build a transliteration system using the Moses toolkit. The transliteration system is applied as a post-processing step to transliterate OOVs.

The morphological reduction of Russian (cf. section 5) does not process most of the OOVs as they are also unknown to the POS tagger. So OOVs that we get are in their original form. When transliterating them, the inflected forms generate wrong English transliterations as inflectional suffixes get transliterated too, specially OOV named entities. We solved this problem by stemming the OOVs based on a list of suffixes (а, ом, ы, е, ой, у) and transliterating the stemmed forms.

Original corpus

SYS	WMT-2012	WMT-2013
GIZA++	32.51	25.5
TA-GIZA++	33.40	25.9*

Morph-reduced

SYS	WMT-2012	WMT-2013
GIZA++	31.22	24.3
TA-GIZA++	31.40	24.45

Table 7: Russian to English machine translation system evaluated on WMT-2012 and WMT-2013. Human evaluation in WMT13 is performed on the system trained using the original corpus with TA-GIZA++ for alignment (marked with *).

Experiments and results We trained the systems separately on GIZA++ and transliteration augmented-GIZA++ (TA-GIZA++) to compare their results; for more details see Sajjad et al. (2013). All systems are tuned using PROv1 (Nakov et al., 2012). The translation output is post-processed to transliterate OOVs.

Table 7 summarizes the results of RU-EN translation systems trained on the original corpus and on the morph-reduced corpus. Using TA-GIZA++ alignment gives the best results for both WMT-2012 and WMT-2013, leading to an improvement of 0.4 BLEU points.

The system built on the morph-reduced data leads to decreased BLEU results. However, the percentage of OOVs is reduced for both test sets when using the morph-reduced data set compared to the original data. An analysis of the output showed that the morph-reduced system makes mistakes in choosing the right tense of the verb, which might be one reason for this outcome. In the future, we would like to investigate this issue in detail.

6 German to English and English to German

We submitted systems for DE-EN and EN-DE which used constituent parses for pre-reordering. For DE-EN we also deal with word formation issues such as compound splitting. We did not perform inflectional normalization or generation for German due to time constraints, instead focusing our efforts on these issues for French and Russian as previously described.

German to English German has a wider diversity of clausal orderings than English, all of which need to be mapped to the English SVO order. This is a difficult problem to solve during inference, as shown for hierarchical SMT by Fabienne Braune and Fraser (2012) and for phrase-based SMT by Bisazza and Federico (2012).

We syntactically parsed all of the source side sentences of the parallel German to English data available, and the tuning, test and blindtest sets. We then applied reordering rules to these parses. We use the rules for reordering German constituent parses of Collins et al. (2005) together with the additional rules described by Fraser (2009). These are applied as a preprocess to all German data.

For parsing the German sentences, we used the generative phrase-structure parser BitPar with optimizations of the grammar, as described by Fraser et al. (2013). The parser was trained on the Tiger Treebank (Brants et al., 2002) along with utilizing the Europarl corpus as unlabeled data. At the training of Bitpar, we followed the targeted self-training approach (Katz-Brown et al., 2011) as follows. We parsed the whole Europarl corpus using a grammar trained on the Tiger corpus and extracted the 100-best parse trees for each sentence. We selected the parse tree among the 100 candidates which got the highest *usefulness scores* for the reordering task. Then we trained a new grammar on the concatenation of the Tiger corpus and the automatic parses from Europarl.

The usefulness score estimates the value of a parse tree for the reordering task. We calculated this score as the similarity between the word order achieved by applying the parse tree-based reordering rules of Fraser (2009) and the word order indicated by the automatic word alignment between the German and English sentences in Europarl. We used the Kendall's Tau Distance as the similarity metric of two word orderings (as suggested by Birch and Osborne (2010)).

Following this, we performed linguistically-informed compound splitting, using the system of Fritzinger and Fraser (2010), which disambiguates competing analyses from the high-recall Stuttgart Morphological Analyzer SMOR (Schmid et al., 2004) using corpus statistics. We also split German portmanteaus like *zum* → *zu dem* (meaning *to the*).

system	BLEU (ci)	BLEU (cs)	system name
DE-EN (OSM)	27.60	26.12	*MES*
DE-EN (OSM) BitPar not self-trained	27.48	25.99	not submitted
DE-EN (Moses)	27.14	25.65	*MES-Szeged-reorder-split*
DE-EN (Moses) BitPar not self-trained	26.82	25.36	not submitted
EN-DE (Moses)	19.68	18.97	*MES-reorder*

Table 8: Results on WMT-2013 (blindtest)

English to German The task of mapping English SVO order to the different clausal orders in German is difficult. For our English to German systems, we solved this by parsing the English and applying the system of Gojun and Fraser (2012) to reorder English into the correct German clausal order (depending on the clause type which is detected using the English parse, see (Gojun and Fraser, 2012) for further details).

We primarily used the Charniak-Johnson generative parser (Charniak and Johnson, 2005) to parse the English Europarl data and the test data. However, due to time constraints we additionally used Berkeley parses of about 400K Europarl sentences and the other English parallel training data. We also left a small amount of the English parallel training data unparsed, which means that it was not reordered. For tune, test and blindtest (WMT-2013), we used the Charniak-Johnson generative parser.

Experiments and results We used all available training data for constrained systems; results for the WMT-2013 set are given in table 8. For the contrastive BitPar results, we reparsed WMT-2013.

7 Conclusion

We presented 5 systems dealing with complex morphology. For two language pairs with a morphologically rich source language (FR and RU), the input was reduced to a simplified representation containing only translation-relevant morphological information (e.g. number on nouns). We also used reordering techniques for DE-EN and EN-DE. For translating into a language with rich morphology (EN-FR), we applied a two-step method that first translates into a stemmed representation of the target language and then generates inflected forms based on morphological features predicted on monolingual data.

Acknowledgments

We would like to thank the anonymous reviewers for their helpful feedback and suggestions, Daniel Quernheim for providing Berkeley parses of some of the English data, Stefan Rüd for help with the manual evalution, and Philipp Koehn and Barry Haddow for providing data and alignments.

Nadir Durrani was funded by the European Union Seventh Framework Programme (FP7/2007-2013) under grant agreement n. 287658. Alexander Fraser was funded by Deutsche Forschungsgemeinschaft grant Models of Morphosyntax for Statistical Machine Translation and from the European Community's Seventh Framework Programme (FP7/2007-2013) under Grant Agreement n. 248005. Marion Weller was funded from the European Community's Seventh Framework Programme (FP7/2007-2013) under Grant Agreement n. 248005. Svetlana Smekalova was funded by Deutsche Forschungsgemeinschaft grant Models of Morphosyntax for Statistical Machine Translation. Helmut Schmid and Max Kisselew were supported by Deutsche Forschungsgemeinschaft grant SFB 732. Richárd Farkas was supported by the European Union and the European Social Fund through project FuturICT.hu (grant n. TÁMOP-4.2.2.C-11/1/KONV-2012-0013). This publication only reflects the authors' views.

References

A. Abeillé, L. Clément, and F. Toussenel. 2003. Building a treebank for french. In A. Abeillé, editor, *Treebanks*. Kluwer, Dordrecht.

Alexandra Birch and Miles Osborne. 2010. Lrscore for evaluating lexical and reordering quality in mt. In *Proceedings of ACL WMT and MetricsMATR*, Uppsala, Sweden.

Arianna Bisazza and Marcello Federico. 2012. Modified distortion matrices for phrase-based statistical machine translation. In *ACL*, pages 478–487.

Sabine Brants, Stefanie Dipper, Silvia Hansen, Wolfgang Lezius, and George Smith. 2002. The TIGER treebank. In *Proceedings of the Workshop on Treebanks and Linguistic Theories*.

Eugene Charniak and Mark Johnson. 2005. Coarse-to-fine n-best parsing and MaxEnt discriminative reranking. In *ACL*, pages 173–180, Ann Arbor, MI, June. Association for Computational Linguistics.

Colin Cherry and George Foster. 2012. Batch tuning strategies for statistical machine translation. In *Proceedings of the North American Chapter of the Association for Computational Linguistics (NAACL)*.

Michael Collins, Philipp Koehn, and Ivona Kučerová. 2005. Clause restructuring for statistical machine translation. In *Porceedings of ACL 2005*.

Nadir Durrani, Helmut Schmid, and Alexander Fraser. 2011. A Joint Sequence Translation Model with Integrated Reordering. In *Proceedings of ACL-HLT 2011*, Portland, Oregon, USA.

Nadir Durrani, Alexander Fraser, and Helmut Schmid. 2013a. Model With Minimal Translation Units, But Decode With Phrases. In *Proceedings of NAACL 2013*, Atlanta, Georgia, USA.

Nadir Durrani, Helmut Schmid, Alexander Fraser, Hassan Sajjad, and Richárd Farkas. 2013b. Munich-Edinburgh-Stuttgart Submissions of OSM Systems at WMT13. In *Proceedings of the Eighth Workshop on Statistical Machine Translation*, Sofia, Bulgaria.

Anita Gojun Fabienne Braune and Alexander Fraser. 2012. Long-distance reordering during search for hierarchical phrase-based SMT. In *Proceedings of EAMT 2012*.

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

Alexander Fraser, Helmut Schmid, Richárd Farkas, Renjing Wang, and Hinrich Schütze. 2013. Knowledge sources for constituent parsing of German, a morphologically rich and less-configurational language. *Computational Linguistics - to appear*.

Alexander Fraser. 2009. Experiments in morphosyntactic processing for translating to and from German. In *EACL WMT*.

Fabienne Fritzinger and Alexander Fraser. 2010. How to avoid burning ducks: Combining linguistic analysis and corpus statistics for German compound processing. In *ACL WMT and Metrics MATR*.

Anita Gojun and Alexander Fraser. 2012. Determining the placement of German verbs in English-to-German SMT. In *Proceedings of EACL 2012*.

Jason Katz-Brown, Slav Petrov, Ryan McDonald, Franz Och, David Talbot, Hiroshi Ichikawa, Masakazu Seno, and Hideto Kazawa. 2011. Training a parser for machine translation reordering. In *Proceedings of EMNLP 2011*, Edinburgh, Scotland.

Thomas Lavergne, Olivier Cappé, and François Yvon. 2010. Practical very large scale CRFs. In *Proceedings of ACL 2010*, pages 504–513.

Preslav Nakov, Francisco Guzmán, and Stephan Vogel. 2012. Optimizing for sentence-level BLEU+1 yields short translations. Mumbai, India.

Hassan Sajjad, Alexander Fraser, and Helmut Schmid. 2011. An algorithm for unsupervised transliteration mining with an application to word alignment. In *Proceedings of ACL 2011*, Portland, USA.

Hassan Sajjad, Alexander Fraser, and Helmut Schmid. 2012. A statistical model for unsupervised and semi-supervised transliteration mining. In *Proceedings of ACL 2012*, Jeju, Korea.

Hassan Sajjad, Svetlana Smekalova, Nadir Durrani, Alexander Fraser, and Helmut Schmid. 2013. QCRI-MES Submission at WMT13: Using Transliteration Mining to Improve Statistical Machine Translation. In *Proceedings of the Eighth Workshop on Statistical Machine Translation*, Sofia, Bulgaria.

Helmut Schmid and Florian Laws. 2008. Estimation of conditional probabilities with decision trees and an application to fine-grained pos tagging. In *Proceedings of COLING 2008*, Stroudsburg, PA, USA.

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

Helmut Schmid. 1994. Probabilistic part-of-speech tagging using decision trees. In *Proceedings of the International Conference on New Methods in Language Processing*.

Holger Schwenk and Philipp Koehn. 2008. Large and diverse language models for statistical machine translation. In *Proceedings of IJCNLP 2008*.

Serge Sharoff, Mikhail Kopotev, Tomaz Erjavec, Anna Feldman, and Dagmar Divjak. 2008. Designing and evaluating russian tagsets. In *Proceedings of LREC 2008*.

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

Zhenxia Zhou. 2007. Entwicklung einer französischen Finite-State-Morphologie. Diploma Thesis, Institute for Natural Language Processing, University of Stuttgart.