Producing Unseen Morphological Variants in Statistical Machine Translation

Matthias Huck[1], **Aleš Tamchyna**[1,2], **Ondřej Bojar**[2] **and Alexander Fraser**[1]

[1]LMU Munich, Munich, Germany
[2]Charles University in Prague, Prague, Czech Republic
{mhuck,fraser}@cis.lmu.de
{tamchyna,bojar}@ufal.mff.cuni.cz

Abstract

Translating into morphologically rich languages is difficult. Although the coverage of lemmas may be reasonable, many morphological variants cannot be learned from the training data. We present a statistical translation system that is able to produce these inflected word forms. Different from most previous work, we do not separate morphological prediction from lexical choice into two consecutive steps. Our approach is novel in that it is integrated in decoding and takes advantage of context information from both the source language and the target language sides.

1 Introduction

Morphologically rich languages exhibit a large amount of inflected word surface forms for most lemmas, which poses difficulties to current statistical machine translation (SMT) technology. SMT systems, such as phrase-based translation (PBT) engines (Koehn et al., 2003), are trained on parallel corpora and can learn the vocabulary that is observed in the data. After training, the decoder can output words which have been seen on the target side of the corpus, but no unseen words.

Sparsity of morphological variants leads to many linguistically valid morphological word forms remaining unseen in practical scenarios. This is a substantial issue under low-resource conditions, but the problem persists even with larger amounts of parallel training data. When translating into the morphologically rich language, the system fails at producing the unseen morphological variants, leading to major translation errors.

Consider the Czech example in Table 1. A small parallel corpus of 50K English-Czech sentences contains only a single variant of the morphological forms of the Czech lemma "čéška" (plural of English: "kneecap"), out of seven syntactically valid cases. The situation improves as we add in more training data (500K/5M/50M), but we can generally not expect the SMT system to learn all variants of each known lemma. In Czech, the number of possible variants is even larger for other word categories such as verbs or adjectives. Adjectives, for instance, have different suffixes depending on case, number, and gender of the governing noun.

case	surface form	50K	500K	5M	50M
1	čéšky	•	•	•	•
2	čéšek	–	•	•	•
3	čéškám	–	–	•	•
4	čéšky	○	○	•	•
5	čéšky	○	○	○	○
6	čéškách	–	•	•	•
7	čéškami	–	–	–	•

Table 1: Morphological variants of the Czech lemma "čéška". For differently sized corpora (50K/500K/5M/50M), "•" indicates that the variant is present, and "○" that the same surface form realization occurs, but in a different syntactic case.

In this paper, we propose an extension to phrase-based SMT that allows the decoder to produce *any* morphological variant of all known lemmas. We design techniques for generating and scoring unseen morphological variants fully integrated into phrase-based search, with the decoder being able to choose freely amongst all possible morphological variants. Empirically, we observe considerable gains in translation quality especially under medium- to low-resource conditions.

2 Related Work

Translation into morphologically rich languages is often tackled through "*two-step*", i.e., separate modules for morphological prediction and generation (Toutanova et al., 2008; Bojar and Kos, 2010;

Fraser et al., 2012; Burlot et al., 2016). An important problem is that lexical choice (of the lemma) is carried out in a separate step from morphological prediction.

Factored machine translation with separate translation and generation models represents a different approach, operating with a single-step search. However, too many options in decoding cause a blow-up of the search space; and useful information is dropped when modeling *source_word→target_lemma* and *target_lemma→target_word* separately.

Word forms not seen in parallel data are sometimes still available in monolingual data. *Back-translation* (Bojar and Tamchyna, 2011) takes advantage of this. The monolingual target language data is lemmatized, automatically translated to the source language, and the translations are aligned with the original, inflected target corpus to produce supplementary training data. Disadvantages are both the computational expense and that the back-translated text may contain errors.

Previous work on *synthetic phrases* by Chahuneau et al. (2013) is most similar to our work. They commit to generation of a single candidate inflection of a lemma prior to decoding, chosen only based on a hierarchical rule and source-side information, a significant limitation. We instead consider all morphological variants, and we are able to use dynamically-generated target-side context in choosing the correct variant, which is critical for capturing phenomena such as target-side verb-subject agreement, or the agreement between a preposition marking case and the case on the noun it marks.

3 Generating Unseen Morphological Variants

We investigate an approach based on synthesized morphological variants. A morphological generation tool is utilized to synthesize all valid morphological forms from target-side lemmas. The phrase table is then augmented with additional entries to provide complete coverage.

We process single target-word entries from the baseline phrase table and feed the lemmatized target word into the morphological generation tool. If its output contains morphological forms that are not known as translations of the source side of the phrase, we add these morphological variants as new translation options. We consider two settings:

feature type	configurations
source indicator	l, t
source internal	l, l+r, l+p, t, r+p
source context	l (-3,3), t (-5,5)
target indicator	l, t
target internal	l, t
target context	l (-2), t (-2)

Table 2: Feature templates for the discriminative classifier: l (lemma), t (morphosyntactic tag), r (syntactic role), p (lemma of dependency parent). Numbers in parentheses indicate context size.

(1.) **word**, where morphological word forms are generated from phrase table entries of length 1 on both source and target side, and (2.) **mtu** (for "minimal translation unit"), where the phrase source side can have arbitrary length.

Morphological generation for Czech, for instance, can be performed with the MorphoDiTa toolkit (Straková et al., 2014), which we will use in our experiments. MorphoDiTa knows a dictionary of most Czech lemmas and can generate all their morphological variants (Hajič, 2004).

When not restricted, the morphological generator also produces forms which do not match in number, tense, degree of comparison, or even negation. This may be undesirable and we therefore define a *tag template*. The tag template prevents the generation of some forms of the given Czech lemma. The template only allows freedom in the following morphological categories: gender, case, person, possessor's number, and possessor's gender. All other attributes must match the original Czech word form. The morphosyntax of the English source is not used to impose further constraints. We will mark this configuration with an asterisk ($\star$) in our experiments.

4 Scoring Unseen Morphological Variants

Assigning dependable model scores to synthesized morphological forms is a primary challenge. During decoding, the artificially added phrase table entries compete with baseline phrases that had been directly extracted from the parallel training data. The correct choice has to be determined in search based on model scores.

A phrase-based model with linguistically motivated *factors* (Koehn and Hoang, 2007) enables us to achieve better generalization capabilities when translating into a morphologically rich language.

system \ newstest	2014 BLEU	2015 BLEU	2016 BLEU
baseline 50K	12.4	10.8	11.8
+ morph-vw-50K	12.2	10.6	11.8
+ synthetic (word)	13.4	11.3	12.5
+ morph-vw-50K	13.4	11.4	12.7
+ synthetic (word⋆)	13.3	11.3	12.5
+ morph-vw-50K	13.3	11.3	12.7
+ synthetic (mtu)	**13.5**	**11.5**	12.7
+ morph-vw-50K	13.4	11.4	12.7
+ synthetic (mtu⋆)	13.4	11.3	12.9
+ morph-vw-50K	**13.5**	**11.5**	**13.1**

Table 3: English→Czech experimental results using 50K training sentence pairs.

system \ newstest	2014 BLEU	2015 BLEU	2016 BLEU
baseline 500K	17.7	14.4	16.1
+ morph-vw-500K	17.6	14.4	16.5
+ synthetic (word)	18.1	14.7	16.4
+ morph-vw-500K	18.4	15.2	17.3
+ synthetic (word⋆)	18.0	14.8	16.6
+ morph-vw-500K	18.2	14.9	17.0
+ synthetic (mtu)	18.1	14.8	16.6
+ morph-vw-500K	18.5	15.3	17.3
+ synthetic (mtu⋆)	18.3	15.0	16.9
+ morph-vw-500K	**18.6**	**15.4**	**17.4**

Table 4: English→Czech experimental results using 500K training sentence pairs.

system \ newstest	2014 BLEU	2015 BLEU	2016 BLEU
baseline 5M	20.8	16.8	18.9
+ morph-vw-5M	20.9	16.8	19.0
+ synthetic (word)	20.9	17.0	19.0
+ morph-vw-5M	**21.1**	17.0	19.0
+ synthetic (word⋆)	20.7	16.8	19.0
+ morph-vw-5M	20.4	16.4	18.7
+ synthetic (mtu)	20.6	**17.2**	19.0
+ morph-vw-5M	21.0	16.9	19.0
+ synthetic (mtu⋆)	20.8	17.1	**19.1**
+ morph-vw-5M	20.9	16.8	19.0

Table 5: English→Czech experimental results using 5M training sentence pairs.

system \ newstest	2014 BLEU	2015 BLEU	2016 BLEU
baseline 50M	22.3	18.1	20.5
+ morph-vw-50M	**22.7**	18.2	20.7
+ synthetic (word)	22.3	18.2	20.5
+ morph-vw-50M	22.3	18.1	20.5
+ synthetic (word⋆)	22.3	18.1	20.4
+ morph-vw-50M	22.5	18.1	20.6
+ synthetic (mtu)	22.3	18.1	20.5
+ morph-vw-50M	**22.7**	**18.3**	**20.8**
+ synthetic (mtu⋆)	22.3	17.9	20.3
+ morph-vw-50M	22.4	18.1	20.5

Table 6: English→Czech experimental results using 50M training sentence pairs.

In our baseline systems, we already draw on lemmas and morphosyntactic tags as factors on the target side, in addition to word surface forms.[1] The additional target-side factors allow us to integrate features that independently model word sense (in terms of the lemma) and morphological attributes (in terms of the morphosyntactic tag). All our translation engines (cf. Section 5) incorporate n-gram LMs over lemmas and over morphosyntactic tags, and an operation sequence model (OSM) (Durrani et al., 2013) with lemmas on the target side. These models counteract sparsity, and where models over surface forms fail for unseen variants, they still assign scores which are based on reliable probability estimates.

When enhancing a system with synthesized phrase table entries, we add further features. Since the usual phrase translation and lexical translation log-probabilities over surface forms cannot be estimated for unseen morphological variants, but all new variants are generated from existing lemmas, we utilize the corresponding log-probabilities over target lemmas. Those can be extracted from the parallel training data and added to the synthesized entries. For baseline phrase table entries, we retain their four baseline phrase translation and lexical translation features, meaning that features over target lemmas score synthesized entries and features over surface forms score baseline entries. The features have separate weights in the model combination. Furthermore, a binary indicator distinguishes baseline phrases from synthesized phrases.

The final key to our approach is using a discriminative classifier (**morph-vw**, *Vowpal Wabbit*[2] *for Morphology*) which can take context from both the source side and the target side into account, as in (Tamchyna et al., 2016). We design feature templates for the classifier that generalize to unseen morphological variants, as listed in Table 2. "Indicator" features are concatenations of words inside

[1] But note that our factored systems operate without a division into separate translation and generation models.

[2] https://hunch.net/~vw/

the phrase, "internal" features represent each word in the phrase separately. Context features on the source side capture a fixed-sized window around the phrase. Target-side context is only to the left of the current phrase. The feature set is designed to force the classifier to learn two independent components: semantic (choosing the right lemma) and morphosyntactic (choosing the right tag, i.e., morphological variant of a word). When scoring an unseen morphological variant of a known word, these two independent components should still be able to assign meaningful scores to the translation. Note that the features require lemmatization and tagging on both sides and a dependency parse of the source side.

5 Empirical Evaluation

For an empirical evaluation of our technique, we build baseline phrase-based SMT engines using `Moses` (Koehn et al., 2007). We then enrich these baselines with linguistically motivated morphological variants that are unseen in the parallel training data, and we augment the model with the discriminative classifier to guide morphological selection during decoding. Different flavors of synthetic morphological variants are compared, each either combined with the discriminative classifier or standalone.

We choose English→Czech as a task that is representative for machine translation from a morphologically underspecified language into a morphologically rich language.

5.1 Experimental Setup

We train a phrase-based translation system with three factors on the target side of the translation model (but no separate generation model). The target factors are the word surface form, lemma, and a morphosyntactic tag. We use the Czech positional tagset (Hajič and Hladká, 1998) which fully describes the word's morphological attributes. On the source side we use only surface forms, except for the discriminative classifier, which includes the features as shown in Table 2.

We employ corpora that have been provided for the English→Czech News translation shared task at WMT16 (Bojar et al., 2016b), including the CzEng parallel corpus (Bojar et al., 2016a). Word alignments are created using `fast_align` (Dyer et al., 2013) and symmetrized. We extract phrases up to a maximum length of 7. The phrase table is pre-pruned by applying a minimum score threshold of 0.0001 on the source-to-target phrase translation probability, and the decoder loads a maximum of 100 best translation options per distinct source side. We use cube pruning in decoding. Pop limit and stack limit for cube pruning are set to 1000 for tuning and to 5000 for testing. The distortion limit is 6. Weights are tuned on newstest2013 with k-best MIRA (Cherry and Foster, 2012) over 200-best lists for 25 iterations. Translation quality is measured in BLEU (Papineni et al., 2002) on three different test sets, newstest2014, newstest2015, and newstest2016.[3]

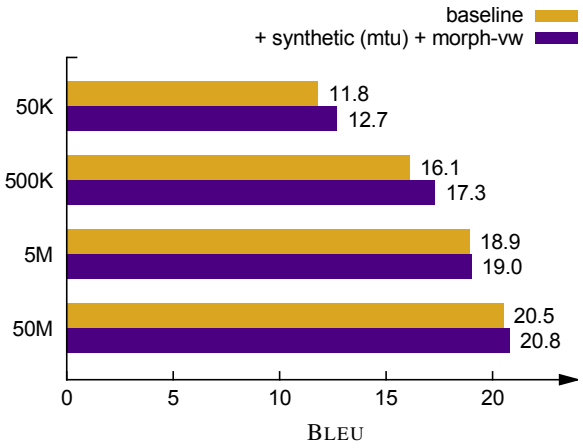

Figure 1: Visualization of the English→Czech translation quality on newstest2016, showing the benefit of our approach under different training resource conditions (50K/500K/5M/50M).

Our training data amounts to around 50 million bilingual sentences overall, but we conduct sets of experiments with systems trained using different fractions of this data (**50K**, **500K**, **5M**, **50M**). Whereas English→Czech has good coverage in terms of training corpora, we simulate low- and medium-resource conditions for the purpose of drawing more general conclusions. Irrespective of this, we utilize the same large LMs in all setups, assuming that proper amounts of target language monolingual data can often be gathered, even when parallel data is scarce. All other models (including the *morph-vw*) are trained using only the fraction of data as chosen for the respective set of experiments, and synthesized phrase table entries with generated morphological variants are produced individually for each baseline phrase table.

[3] We evaluate case-sensitive with `mteval-v13a.pl -c`, comparing post-processed hypotheses against the raw reference.

input: now , six in 10 Republicans have a favorable view of Donald Trump .

baseline: teď , šest **v** 10 **republikáni** mají příznivý **výhled Donald Trump** .
now, six ***in***$_{location}$ *10* ***Republicans***$_{nom}$ *have a_favorable* ***outlook Donald***$_{nom}$ ***Trump***$_{nom}$.

+ synthetic (mtu) + morph-vw: teď , šest **do** deseti republikánů má příznivý názor na Donalda Trumpa .
now, six **into** *ten*$_{gen}$ *Republicans*$_{gen}$ *have a_favorable opinion of Donald*$_{acc}$ *Trump*$_{acc}$.

Figure 2: Example outputs of 500K system variants. Each translation has a corresponding gloss in italics. Errors are marked in bold. Synthetic phrase translations are underlined.

5.2 Experimental Results and Analysis

Translation results are reported in Tables 3 to 6. Our method is effective at improving BLEU especially in the low- and medium-resource settings, but shows only slight gains in the 5M and 50M scenarios. Overall, *mtu* leads to better results than *word*. When we also add translations to phrases with multiple input words, we give the system more leeway in phrasal segmentation and our synthetic phrases can perhaps be applied more easily.

In the 50K and 500K settings, we obtain considerable improvements even without using the discriminative model. This suggests that our scoring scheme based on lemmas is indeed effective for the synthetic phrase pairs. Additionally, model features such as the OSM with target-side lemmas as well as the LMs over lemmas and over morphosyntactic tags seem to cope with the synthetic word forms reasonably well. However, when we do use the classifier, we obtain a small but consistent further improvement.

Figure 1 visualizes the BLEU scores achieved under the four training resource conditions with the baseline system and with the system extended via synthesized morphological word forms (in the *mtu* variant) plus the discriminative classifier, respectively.

In order to better understand why the improvements fall off as we increase training data size, we measure target-side out-of-vocabulary (OOV) rates of the various settings. Our aim is to quantify the potential improvement that our method can bring. Table 7 shows the statistics: at 50K, the baseline OOV rate is nearly 17 % and our technique successfully reduces it to less than 10 %. The relative reduction of the OOV rate is quite steady as training data size increases.

Figure 2 illustrates the effect of our technique in a medium-size setting (500K). The baseline system is forced to use the incorrect nominative case due to the lack of required surface forms. Our method provides these inflections ("republikánů", "Trumpa") and produces a mostly grammatical translation (but is still unable to correctly translate the preposition "in").

setup	#phrases full	#phrases filtered	OOV (target) types	OOV (target) tokens
baseline 50K	1.6 M	0.2 M	45.8 %	16.6 %
+ synthetic (word)	7.8 M	3.9 M	26.7 %	9.9 %
+ synthetic (word⋆)	2.1 M	0.5 M	35.0 %	12.5 %
+ synthetic (mtu)	19.0 M	5.7 M	26.2 %	9.7 %
+ synthetic (mtu⋆)	3.0 M	0.7 M	34.5 %	12.3 %
baseline 500K	14.5 M	1.4 M	21.0 %	7.1 %
+ synthetic (word)	44.3 M	16.0 M	11.9 %	4.2 %
+ synthetic (word⋆)	16.9 M	2.5 M	15.2 %	5.2 %
+ synthetic (mtu)	134.4 M	25.8 M	11.6 %	4.1 %
+ synthetic (mtu⋆)	24.0 M	3.3 M	14.9 %	5.1 %
baseline 5M	126.6 M	7.4 M	9.1 %	3.1 %
+ synthetic (word)	254.4 M	58.0 M	5.8 %	2.2 %
+ synthetic (word⋆)	137.1 M	11.4 M	6.7 %	2.4 %
+ synthetic (mtu)	953.3 M	105.9 M	5.7 %	2.1 %
+ synthetic (mtu⋆)	192.1 M	15.0 M	6.6 %	2.4 %
baseline 50M	996.5 M	23.4 M	4.9 %	1.7 %
+ synthetic (word)	1 415.2 M	122.2 M	3.6 %	1.3 %
+ synthetic (word⋆)	1 030.7 M	30.4 M	4.0 %	1.4 %
+ synthetic (mtu)	6 256.2 M	287.4 M	3.5 %	1.3 %
+ synthetic (mtu⋆)	1 414.1 M	42 6 M	3.9 %	1.4 %

Table 7: Phrase table statistics. We report sizes of the full phrase tables as well as after filtering towards the newstest2016 source. Target-side OOV rates are calculated by comparing newstest2016 references against the filtered phrase tables.

6 Conclusion

We have studied the important problem of modeling all morphological variants of our SMT system's vocabulary. We showed that we can augment our system's vocabulary with the missing variants and that we can effectively score these variants using a discriminative lexicon utilizing both source and target context. We have shown that this leads to substantial BLEU score improvements, particularly on small to medium resource translation tasks. Given the limited training data available for translation to many morphologically rich languages, our approach is widely applicable.

Acknowledgments

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreements № 644402 (*HimL*) and № 645452 (*QT21*), from the European Research Council (ERC) under grant agreement № 640550, and from the DFG grant *Models of Morphosyntax for Statistical Machine Translation (Phase Two)*. This work has been using language resources and tools developed and distributed by the LINDAT/CLARIN project of the Ministry of Education, Youth and Sports of the Czech Republic (project LM2015071).

References

Ondřej Bojar and Kamil Kos. 2010. 2010 Failures in English-Czech Phrase-Based MT. In *Proceedings of the Joint Fifth Workshop on Statistical Machine Translation and MetricsMATR*, pages 60–66, Uppsala, Sweden, July. Association for Computational Linguistics.

Ondřej Bojar and Aleš Tamchyna. 2011. Improving Translation Model by Monolingual Data. In *Proceedings of the Sixth Workshop on Statistical Machine Translation*, pages 330–336, Edinburgh, Scotland, July. Association for Computational Linguistics.

Ondřej Bojar, Ondřej Dušek, Tom Kocmi, Jindřich Libovický, Michal Novák, Martin Popel, Roman Sudarikov, and Dušan Variš, 2016a. *CzEng 1.6: Enlarged Czech-English Parallel Corpus with Processing Tools Dockered*, pages 231–238. Springer International Publishing, Cham.

Ondřej Bojar, Rajen Chatterjee, Christian Federmann, Yvette Graham, Barry Haddow, Matthias Huck, Antonio Jimeno Yepes, Philipp Koehn, Varvara Logacheva, Christof Monz, Matteo Negri, Aurelie Neveol, Mariana Neves, Martin Popel, Matt Post, Raphael Rubino, Carolina Scarton, Lucia Specia, Marco Turchi, Karin Verspoor, and Marcos Zampieri. 2016b. Findings of the 2016 Conference on Machine Translation. In *Proceedings of the First Conference on Machine Translation*, pages 131–198, Berlin, Germany, August. Association for Computational Linguistics.

Franck Burlot, Elena Knyazeva, Thomas Lavergne, and François Yvon. 2016. Two-Step MT: Predicting Target Morphology. In *Proceedings of the International Workshop on Spoken Language Translation (IWSLT)*, Seattle, Washington, USA, December.

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*, pages 1677–1687, Seattle, Washington, USA, October. Association for Computational Linguistics.

Colin Cherry and George Foster. 2012. Batch Tuning Strategies for Statistical Machine Translation. In *Proceedings of the 2012 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 427–436, Montréal, Canada, June. Association for Computational Linguistics.

Nadir Durrani, Alexander Fraser, and Helmut Schmid. 2013. Model With Minimal Translation Units, But Decode With Phrases. In *Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 1–11, Atlanta, Georgia, June. Association for Computational Linguistics.

Chris Dyer, Victor Chahuneau, and Noah A. Smith. 2013. A Simple, Fast, and Effective Reparameterization of IBM Model 2. In *Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 644–648, Atlanta, Georgia, June. Association for Computational Linguistics.

Alexander Fraser, Marion Weller, Aoife Cahill, and Fabienne Cap. 2012. Modeling Inflection and Word-Formation in SMT. In *Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics*, pages 664–674, Avignon, France, April. Association for Computational Linguistics.

Jan Hajič and Barbora Hladká. 1998. Tagging Inflective Languages: Prediction of Morphological Categories for a Rich Structured Tagset. In *Proceedings of the 36th Annual Meeting of the Association for Computational Linguistics and 17th International Conference on Computational Linguistics, Volume 1*, pages 483–490, Montreal, Quebec, Canada, August. Association for Computational Linguistics.

Jan Hajič. 2004. *Disambiguation of Rich Inflection (Computational Morphology of Czech)*. Karolinum, Charles University Press, Prague, Czech Republic.

Philipp Koehn and Hieu Hoang. 2007. Factored Translation Models. In *Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL)*, pages 868–876, Prague, Czech Republic, June. Association for Computational Linguistics.

Philipp Koehn, Franz Josef Och, and Daniel Marcu. 2003. Statistical Phrase-Based Translation. In *Proceedings of the 2003 Human Language Technology*

Conference of the North American Chapter of the Association for Computational Linguistics, pages 127–133, Edmonton, Canada, May/June. Association for Computational Linguistics.

Philipp Koehn, Hieu Hoang, Alexandra Birch, Chris Callison-Burch, Marcello Federico, Nicola Bertoldi, Brooke Cowan, Wade Shen, Christine Moran, Richard Zens, Chris Dyer, Ondrej Bojar, Alexandra Constantin, and Evan Herbst. 2007. Moses: Open Source Toolkit for Statistical Machine Translation. In *Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics Companion Volume Proceedings of the Demo and Poster Sessions*, pages 177–180, Prague, Czech Republic, June. Association for Computational Linguistics.

Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. 2002. Bleu: a Method for Automatic Evaluation of Machine Translation. In *Proceedings of 40th Annual Meeting of the Association for Computational Linguistics*, pages 311–318, Philadelphia, Pennsylvania, USA, July. Association for Computational Linguistics.

Jana Straková, Milan Straka, and Jan Hajič. 2014. Open-Source Tools for Morphology, Lemmatization, POS Tagging and Named Entity Recognition. In *Proceedings of 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, pages 13–18, Baltimore, Maryland, June. Association for Computational Linguistics.

Aleš Tamchyna, Alexander Fraser, Ondřej Bojar, and Marcin Junczys-Dowmunt. 2016. Target-Side Context for Discriminative Models in Statistical Machine Translation. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1704–1714, Berlin, Germany, August. Association for Computational Linguistics.

Kristina Toutanova, Hisami Suzuki, and Achim Ruopp. 2008. Applying Morphology Generation Models to Machine Translation. In *Proceedings of ACL-08: HLT*, pages 514–522, Columbus, Ohio, June. Association for Computational Linguistics.