LMU Munich's Neural Machine Translation Systems for News Articles and Health Information Texts

Matthias Huck, Fabienne Braune, Alexander Fraser

Center for Information and Language Processing
LMU Munich
Munich, Germany

`{mhuck,braune,fraser}@cis.lmu.de`

Abstract

This paper describes the LMU Munich English→German machine translation systems. We participated with neural translation engines in the WMT17 shared task on machine translation of news, as well as in the biomedical translation task. LMU Munich's systems deliver competitive machine translation quality on both news articles and health information texts.

1 Introduction

The Center for Information and Language Processing at LMU Munich has a strong track record at building statistical machine translation (SMT) systems for various language pairs, e.g. for translation between English and German, Czech, Romanian, Russian, or French. LMU has frequently participated in WMT machine translation shared tasks in recent years (Bojar et al., 2016, 2015, 2014, 2013), competing (and also collaborating) internationally in an open evaluation campaign with other leading research labs from both academia and industry.

Research on various different types of machine translation models has previously been conducted at LMU. Core SMT paradigms for LMU's past shared task participations include phrase-based models (Cap et al., 2015, 2014b; Weller et al., 2013; Sajjad et al., 2013), hierarchical phrase-based models (Huck et al., 2016; Peter et al., 2016), operation sequence models (Durrani et al., 2013), and hybrids of statistical approaches with rule-based and deep syntactic components (Tamchyna et al., 2016b).

At this year's *EMNLP 2017 Second Conference on Machine Translation (WMT17),*[1] LMU participated in two shared tasks: the shared task on *machine translation of news* and the *biomedical translation task*. We submitted the output of our English→German machine translation systems. The system for the news task was trained under "constrained" conditions, employing only permissible resources as defined by the shared task organizers. The system for the biomedical task builds upon our news task system, but was domain-adapted towards the medical domain via the usage of additional parallel training data from the in-domain sections of the UFAL Medical Corpus v.1.0.

We have trained *neural machine translation* (NMT) models this year. Neural network models for machine translation (Sutskever et al., 2014; Bahdanau et al., 2014) are now largely successful for many language pairs and domains. This has for instance become apparent with the University of Edinburgh's excellent results in the WMT16 news translation shared task with neural systems (Sennrich et al., 2016a), which outperformed most other submitted systems, including Edinburgh's own traditional SMT engines (Williams et al., 2016). LMU's English→German neural machine translation systems confirm this trend. We have achieved competitive performance—in terms of translation quality as measured with BLEU (Papineni et al., 2002)—in both shared tasks that we participated in.[2]

A unique characteristic of the LMU English→German NMT systems is a linguistically informed, cascaded word segmentation technique that we developed and applied to the German target language side of the training data. Amongst other aspects, SMT research at LMU is

---

[2] Our LMU Munich primary system is ranked second in BLEU on the submission website, http://matrix.statmt.org/matrix/systems_list/1869, being outpaced by Edinburgh's WMT17 NMT setup only. In the human evaluation the LMU Munich primary system is ranked first (Bojar et al., 2017).

focusing on investigating linguistically informed methods that improve machine translation into target languages which exhibit a more complex morphosyntax than English (Huck et al., 2017b; Tamchyna et al., 2016a; Ramm and Fraser, 2016; Weller-Di Marco et al., 2016; Braune et al., 2015; Cap et al., 2014a; Fraser et al., 2012). We are taking advantage of our group's longstanding experience regarding handling of complex morphosyntax in SMT, now enriching NMT with novel techniques that specifically tackle target-side morphosyntax.

In the following section of this paper (Section 2), we sketch our linguistically motivated target word segmentation technique. Then we describe how we trained and configured our neural machine translation systems (Section 3). Before concluding the paper, we present empirical results on the two translation tasks, which involve machine translation of news articles and of health information texts (Section 4).

2 Target-side Word Segmentation

Compounding and morphological variation are ubiquitous in the German language and have traditionally been challenging for machine translation into German. We believe that specifically targeting complex morphosyntactic phenomena in the output language is not only essential in traditional phrase-based machine translation, but keeps being valuable in NMT. Most previous work in NMT has focused on linguistically agnostic subword splitting, typically with the primary rationale of limiting the vocabulary size, which is required in NMT for efficiency considerations.

LMU is utilizing a more linguistically-informed target word segmentation approach. By doing so, we hope to achieve three major goals: better vocabulary reduction; reduction of data sparsity; and better open vocabulary translation.

We cascade three different word splitting methods on the German target side.

1. First we apply a suffix splitter that separates common German morphological suffixes from the word stems. We modified the German Snowball stemming algorithm from NLTK[3] for that purpose. Rather than stripping suffixes, our modified code splits them

off. It otherwise behaves just like the Snowball stemming algorithm.

2. Next we apply the empirical compound splitter as described by Koehn and Knight (2003) and as implemented in the Perl script which is part of the Moses toolkit (Koehn et al., 2007). We choose a fairly aggressive configuration of the compound splitter[4] in order to reduce the vocabulary size more than with its parameters as typically chosen for previous phrase-based translation setups in which German compound splitting was used.

3. Since the vocabulary size is still a bit large after suffix splitting and compound splitting, we adopt segmentation using the Byte Pair Encoding (BPE) technique (Gage, 1994; Sennrich et al., 2016c) on top of the other two word splitters. This last step is performed only for efficiency reasons in NMT. Without BPE, the vocabulary size is still almost 100K. We preferred something around 50K, which is more tractable in practice. Suffix splitting and compound splitting alone are not suitable for arbitrary reduction of the vocabulary size. However, we believe that they are more adequate word segmentation techniques than BPE is. So we prefer to split with those linguistically motivated methods, as far as practicable.

Special marker symbols allow us to revert the segmentation in postprocessing. We also introduce a case marker that is placed before any compound-split word in order to restore upper and lower casing, respectively, since the compound splitting approach modifies the casing of compound parts to the version of each part that (standalone) appears most frequently in the corpus.

We were running a comprehensive series of experiments with different target word segmentation strategies on Europarl data beforehand, and we found our cascaded word segmentation to perform clearly better than using BPE only. We furthermore tried prefix splitting, but the results looked less encouraging. Our Europarl results also suggested that the suffix splitting contributes more to improvements in translation quality than the compound splitting does. Huck et al. (2017a) provides further details. For our WMT17 shared task systems, we eventually decided to apply both suffix

---

[3] http://www.nltk.org/_modules/nltk/stem/snowball.html

[4] -min-size 4 -min-count 2 -max-count 999999999

splitting and compound splitting, but to omit pre-fix splitting.

The English source side is simply BPE-segmented.

3 Neural Translation System Setup

We utilize the Nematus implementation (Sennrich et al., 2017) to build encoder-decoder NMT systems with attention and gated recurrent units. We configure dimensions of 500 for the embeddings and 1024 for the hidden layer. We train with the Adam optimizer (Kingma and Ba, 2015), a learning rate of 0.0001, batch size of 50, and dropout with probability 0.2 applied to the hidden layer, but not to source, target, and embeddings. We validate every 10 000 updates and do early stopping when the validation cost has not decreased over ten consecutive control points.

Our initial baseline NMT system is trained using only data from the Europarl corpus (Koehn, 2005) and no other resources, with the Europarl `test2006` set used for validation. We tokenize and frequent-case the data with the standard scripts from the Moses toolkit (Koehn et al., 2007). For our Europarl-trained baseline, sentences of length >50 after tokenization are excluded from the training corpus, all other sentences (1.7 M) are kept in training.

The German compound split model and BPE merge operations are extracted from the Europarl data. In our cascaded word segmentation pipeline, the compound split model is extracted from the training data only after suffix splitting has been applied. Similarly, the BPE operations are extracted after suffix splitting and compound splitting have been applied to the German side of the training corpus. We set the amount of merge operations for BPE to 50K. On the English source side, we apply BPE separately, also with 50K merge operations.

3.1 News Translation Task

For the shared task on machine translation of news (Bojar et al., 2017), we successively improved our initial baseline by incrementally applying the following steps:

1. Adding the News Commentary (NC) and Common Crawl (CC) parallel training data as provided for WMT17 by the organizers of the news translation shared task. We initialize the optimization on the larger corpus with the Europarl-trained baseline model.

2. Adding synthetic training data. The use of automatically translated monolingual data as a supplementary training resource has proved to be effective in SMT for phrase-based, hierarchical, and neural systems (Ueffing et al., 2007; Lambert et al., 2011; Huck et al., 2011; Huck and Ney, 2012; Sennrich et al., 2016b). Sennrich et al. have publicly shared their backtranslations of monolingual WMT News Crawl corpora, which they created for their WMT16 participation (Sennrich et al., 2016a). We exploit the full amount of backtranslations of German data into English.[5] We concatenate the synthetic data and the human-generated parallel training data (Europarl + NC + CC). The optimization is initialized with the pre-trained model from the preceding step.

3. Fine-tuning towards the domain of news articles. We employ the `newstest` development sets from the years 2008 to 2014 as a training corpus. We reduce the learning rate to 0.000001, initialize with the pre-trained model from the preceding step, and optimize on only the small `Devsets2008-14` corpus.

4. Right-to-left reranking. We rerank an n-best list from the system in the preceding step with a right-to-left (r2l) model, where the order of the target sequence is reversed. Liu et al. (2016) have proposed right-to-left reranking for NMT. Earlier work by Freitag et al. (2013) had already established that reverse word order models can be beneficial in phrase-based and hierarchical phrase-based translation. Freitag et al. (2013) utilized reverse word order models by means of a system combination framework (Freitag et al., 2014), though.

Validation is done on `newstest2015` for each of the extended setups. The preprocessing pipeline is not altered when more training data is appended. Particularly, we keep applying the compound split model and BPE operations that have been extracted from only the Europarl corpus, and keep sticking to the vocabulary from Europarl. We force the system to suppress UNK tokens in inference at test time.

---

[5] http://data.statmt.org/rsennrich/wmt16_backtranslations/en-de/

system / newstest	2015 BLEU	2016 BLEU	2017 BLEU
baseline, Europarl-trained	19.4	22.8	18.3
+ NC & CC corpora	26.0	30.2	24.5
+ synthetic data	27.8	32.3	26.1
+ fine-tuning on Devsets2008-14	28.2	32.3	26.6
+ r2l reranking	**28.6**	**33.4**	**27.1**

Table 1: English→German translation results on newstest sets (case-sensitive BLEU). Extensions are applied incrementally.

system / devtest	Cochrane BLEU	NHS24 BLEU
fine-tuning on Devsets2008-14	29.1	26.3
fine-tuning on Medical	35.0	29.5
+ r2l reranking	**35.8**	**30.3**

Table 2: English→German translation results on HimL biomedical sets (case-sensitive BLEU).

3.2 Biomedical Translation Task

For the biomedical translation task (Yepes et al., 2017), we started off with the pre-trained NMT model after step 2 of our news task system engineering and applied the following steps:

1. Fine-tuning towards the domain of health information texts. We employ the in-domain sections of the UFAL Medical Corpus v.1.0 as a training corpus.[6] We set the learning rate to 0.00001, initialize with the pre-trained model, and optimize on only the in-domain medical data.
2. Right-to-left reranking. An ensemble of domain-adapted r2l models worked best.

The HimL (Haddow et al., 2017) tuning sets are used for validation, and we tested separately on the Cochrane and NHS24 parts of the HimL devtest set.[7]

4 Empirical Results

We evaluate case-sensitive with BLEU (Papineni et al., 2002), computed over postprocessed hypotheses against the raw references with mteval-v13a. The results are reported in Table 1 for the news translation task and in Table 2 for the biomedical translation task.

In the news translation task, the Europarl-trained baseline does not get close to state-of-the-art performance on newstest sets. However,

this seems to be mostly due to a domain mismatch (Huck et al., 2015). Once we add in the News Commentary and Common Crawl parallel data, we are able to massively improve the translation quality, by around six to seven BLEU points. Synthetic data gives us a boost of about another two BLEU points. After fine-tuning on Devsets2008-14 towards news articles, we observe a further gain of 0.4 BLEU on newstest2015 but no gain on newstest2016. Reranking with a right-to-left model is effective on all test sets again, with improvements in the range of 0.4 to 1.1 BLEU.

Two LMU submissions have been judged by humans in the manual evaluation for the WMT17 news translation task (Bojar et al., 2017): the output of our final setup with r2l reranking (as a primary submission; "*LMU-nmt-reranked*"), and the single system output without reranking (as a contrastive submission; "*LMU-nmt-single*"). Our primary submission is placed first amongst all evaluated systems. We conjecture that our linguistically-informed target word segmentation approach has contributed to a positive assessment by human evaluators. Interestingly, the contrastive submission was rated significantly worse, affirming the utility of r2l reranking.

A few example translations from our primary submission for the news task are shown in Table 3.

For the translation of health information texts, it is again crucial to adapt the NMT system to the domain. When applying the engine fine-tuned on out-of-domain news data (Devsets2008-14) to Cochrane and NHS24 devtest sets, we see quite a gap as compared to fine-tuning on the in-domain sections of the UFAL Medical Corpus. Right-to-left reranking improves the results by 0.8 BLEU for the biomedical task.

5 Conclusion

LMU Munich has participated with English→German neural machine translation systems in the WMT17 shared tasks on machine translation of news and of biomedical texts. A distinctive feature of LMU's NMT systems is a linguistically informed, cascaded target word segmentation approach. The LMU systems are very competitive in terms of translation quality, achieving top ranks amongst the participants in both tasks. Of all English→German systems manually evaluated in the news task, LMU's primary submission has received the highest human judgment scores.

---

[6]https://ufal.mff.cuni.cz/ufal_medical_corpus

[7]http://www.himl.eu/test-sets

source (preproc.)	the Kurdish community in Germany is expecting tens of thousands of people to arrive at short notice in search of protection , fleeing from Turkey to Germany .
LMU-nmt (plain)	die kurdisch $$e Gemeind $$e in #U deutsch @@ Land rechnet damit , dass zehntaus $$end $$e Mensch $$en #L kurz @@ Frist $$ig auf der Such $$e nach dem Schutz eintreff $$en , der aus der Türkei nach #U deutsch @@ Land gefloh $$en ist .
LMU-nmt (postproc.)	Die kurdische Gemeinde in Deutschland rechnet damit, dass zehntausende Menschen kurzfristig auf der Suche nach dem Schutz eintreffen, der aus der Türkei nach Deutschland geflohen ist.
reference	Die Kurdische Gemeinde Deutschland rechnet kurzfristig mit zehntausenden Schutzsuchenden, die aus der Türkei nach Deutschland flüchten.
source (preproc.)	the situation only worsened over the past year when the world ' biggest producer , China , dumped steel into the global market en masse as a result of weakening domestic demand .
LMU-nmt (plain)	die Lag $$e verschlechtert $$e sich nur im vergang $$en $$en Jahr , als der #L Welt @@ größt $$e Produzent , China , infolg $$e der schwä ## chelnd $$en #U binn @en@ nach @@ Frag $$e #L Mass @en@ Haft Stahl in den global $$en Markt geworf $$en hat .
LMU-nmt (postproc.)	Die Lage verschlechterte sich nur im vergangenen Jahr, als der weltgrößte Produzent, China, infolge der schwächelnden Binnennachfrage massenhaft Stahl in den globalen Markt geworfen hat.
reference	Im vergangenen Jahr verschärfte sich die Lage weiter, als das weltgrößte Erzeugerland China angesichts der schwächelnden heimischen Nachfrage massenhaft Stahl auf den Weltmarkt warf.
source (preproc.)	analysts fear that separatist groups that had been more or less vanquished in recent years , like the Oro ## mo Liberation Front or the Og ## aden National Liberation Front , may try to exploit the turbulence and rearm .
LMU-nmt (plain)	Analyst $$en befürcht $$en , dass separatist $$isch $$e Grupp $$en , die in den letzt $$en Jahr $$en mehr oder wenig $$er bezwung $$en word $$en war $$en , wie die Oro ## mo @-@ #U Befreiung @s@ Front oder der O ## gad $$en National $$e #U Befreiung @s@ Front , versuch $$en könnt $$en , die Turbulenz $$en und die Aufrüst $$ung auszunutz $$en .
LMU-nmt (postproc.)	Analysten befürchten, dass separatistische Gruppen, die in den letzten Jahren mehr oder weniger bezwungen worden waren, wie die Oromo-Befreiungsfront oder der Ogaden Nationale Befreiungsfront, versuchen könnten, die Turbulenzen und die Aufrüstung auszunutzen.
reference	Analytiker befürchten, dass Separatisten wie die Oromo-Befreiungsfront oder die Nationale Befreiungsfront des Ogaden, die in den letzten Jahren mehr oder weniger bezwungen wurden, die Turbulenzen ausnützen und sich wieder bewaffnen könnten.
source (preproc.)	these cele ## bri ## ties are not relatives of famous people , or reality stars , or kids these days who know how to make a good S ## n ## ap ## chat video (although Jen ## ner is all of these things) .
LMU-nmt (plain)	dies $$e Pro ## minent $$en sind kein $$e Verwandt $$en berühmt $$er Mensch $$en oder Re ## ality @-@ Star $$s oder Kind $$er dies $$er Tag $$e , die wiss $$en , wie man ein gut $$es Sna ## p ## ch ## at @-@ Video mach $$en kann (obwohl J ## enn $$er all dies $$e Ding $$e ist) .
LMU-nmt (postproc.)	Diese Prominenten sind keine Verwandten berühmter Menschen oder Reality-Stars oder Kinder dieser Tage, die wissen, wie man ein gutes Snapchat-Video machen kann (obwohl Jenner all diese Dinge ist).
reference	Diese Berühmtheiten sind nicht mit berühmten Personen verwandt oder Reality Stars oder Jugendliche von heute, die wissen, wie man ein gutes Snapchat-Video dreht (auch wenn davon alles auf Jenner zutrifft).
source (preproc.)	the specialists from the 3 ## 4th police inspectorate were able to prove that the thieves , who had travelled to Germany to commit the crimes , had committed four crimes .
LMU-nmt (plain)	die Spezialist $$en der 34. #U Polizei @@ Inspektion konnt $$en beweis $$en , dass die Dieb $$e , die nach #U deutsch @@ Land gereist war $$en , um die Verbrech $$en zu begeh $$en , vier Straftat $$en begang $$en hatt $$en .
LMU-nmt (postproc.)	Die Spezialisten der 34. Polizeiinspektion konnten beweisen, dass die Diebe, die nach Deutschland gereist waren, um die Verbrechen zu begehen, vier Straftaten begangen hatten.
reference	Die Spezialisten des Kriminalkommissariats 34 können den Dieben, die eigens zur Begehung von Straftaten nach Deutschland eingereist waren, vier Taten nachweisen.

Table 3: Example translations, produced with LMU Munich's primary machine translation system for the news task. The table shows the preprocessed English source, the plain system output, the postprocessed system output, and the German reference translation for every 479th sentence from the `newstest2017` evaluation set (excluding the very first of them, sentence 479, since it is too short to be interesting). ## is a BPE split-point, $$en is the suffix *en*, #U and #L are upper and lower case indicators for the first word of compounds, @@ indicates a compound merge-point, @s@ indicates a compound merged with the letter *s* between the parts, etc.

Acknowledgments

References

Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. 2014. Neural Machine Translation by Jointly Learning to Align and Translate. *arXiv e-prints*, abs/1409.0473. Presented at ICLR 2015.

Ondřej Bojar, Christian Buck, Chris Callison-Burch, Christian Federmann, Barry Haddow, Philipp Koehn, Christof Monz, Matt Post, Radu Soricut, and Lucia Specia. 2013. Findings of the 2013 Workshop on Statistical Machine Translation. In *Proceedings of the Eighth Workshop on Statistical Machine Translation*, pages 1–44, Sofia, Bulgaria. Association for Computational Linguistics.

Ondřej Bojar, Christian Buck, Christian Federmann, Barry Haddow, Philipp Koehn, Johannes Leveling, Christof Monz, Pavel Pecina, Matt Post, Herve Saint-Amand, Radu Soricut, Lucia Specia, and Aleš Tamchyna. 2014. Findings of the 2014 Workshop on Statistical Machine Translation. In *Proceedings of the Ninth Workshop on Statistical Machine Translation*, pages 12–58, Baltimore, MD, USA. Association for Computational Linguistics.

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. 2016. Findings of the 2016 Conference on Machine Translation. In *Proceedings of the First Conference on Machine Translation*, pages 131–198, Berlin, Germany. Association for Computational Linguistics.

Ondřej Bojar, Rajen Chatterjee, Christian Federmann, Barry Haddow, Matthias Huck, Chris Hokamp, Philipp Koehn, Varvara Logacheva, Christof Monz, Matteo Negri, Matt Post, Carolina Scarton, Lucia Specia, and Marco Turchi. 2015. Findings of the 2015 Workshop on Statistical Machine Translation. In *Proceedings of the Tenth Workshop on Statistical Machine Translation*, pages 1–46, Lisbon, Portugal. Association for Computational Linguistics.

Ondřej Bojar et al. 2017. Findings of the 2017 Conference on Machine Translation. In *Proceedings of the Second Conference on Machine Translation*, Copenhagen, Denmark. Association for Computational Linguistics.

Fabienne Braune, Nina Seemann, and Alexander Fraser. 2015. Rule Selection with Soft Syntactic Features for String-to-Tree Statistical Machine Translation. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, pages 1095–1101, Lisbon, Portugal. Association for Computational Linguistics.

Fabienne Cap, Alexander Fraser, Marion Weller, and Aoife Cahill. 2014a. How to Produce Unseen Teddy Bears: Improved Morphological Processing of Compounds in SMT. In *Proceedings of the 14th Conference of the European Chapter of the Association for Computational Linguistics*, pages 579–587, Gothenburg, Sweden. Association for Computational Linguistics.

Fabienne Cap, Marion Weller, Anita Ramm, and Alexander Fraser. 2014b. CimS – The CIS and IMS joint submission to WMT 2014 translating from English into German. In *Proceedings of the Ninth Workshop on Statistical Machine Translation*, pages 71–78, Baltimore, MD, USA. Association for Computational Linguistics.

Fabienne Cap, Marion Weller, Anita Ramm, and Alexander Fraser. 2015. CimS - The CIS and IMS Joint Submission to WMT 2015 addressing morphological and syntactic differences in English to German SMT. In *Proceedings of the Tenth Workshop on Statistical Machine Translation*, pages 84–91, Lisbon, Portugal. Association for Computational Linguistics.

Nadir Durrani, Alexander Fraser, Helmut Schmid, Hassan Sajjad, and Richárd Farkas. 2013. Munich-Edinburgh-Stuttgart Submissions of OSM Systems at WMT13. In *Proceedings of the Eighth Workshop on Statistical Machine Translation*, pages 122–127, Sofia, Bulgaria. 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. Association for Computational Linguistics.

Markus Freitag, Minwei Feng, Matthias Huck, Stephan Peitz, and Hermann Ney. 2013. Reverse Word Order Models. In *Proceedings of the XIV Machine Translation Summit*, pages 159–166, Nice, France.

Markus Freitag, Matthias Huck, and Hermann Ney. 2014. Jane: Open Source Machine Translation System Combination. In *Proceedings of the Demonstrations at the 14th Conference of the European Chapter of the Association for Computational Linguistics*, pages 29–32, Gothenburg, Sweden. Association for Computational Linguistics.

Philip Gage. 1994. A New Algorithm for Data Compression. *C Users J.*, 12(2):23–38.

Barry Haddow, Alexandra Birch, Ondřej Bojar, Fabienne Braune, Colin Davenport, Alex Fraser, Matthias Huck, Michal Kašpar, Květoslava Kovaříková, Josef Plch, Anita Ramm, Juliane Ried, James Sheary, Aleš Tamchyna, Dušan Variš, Marion Weller, and Phil Williams. 2017. HimL: Health in my Language. In *Proceedings of the EAMT 2017 User Studies and Project/Product Descriptions*, page 33, Prague, Czech Republic.

Matthias Huck, Alexandra Birch, and Barry Haddow. 2015. Mixed-Domain vs. Multi-Domain Statistical Machine Translation. In *Proceedings of MT Summit XV, vol.1: MT Researchers' Track*, pages 240–255, Miami, FL, USA.

Matthias Huck, Alexander Fraser, and Barry Haddow. 2016. The Edinburgh/LMU Hierarchical Machine Translation System for WMT 2016. In *Proceedings of the First Conference on Machine Translation*, pages 311–318, Berlin, Germany. Association for Computational Linguistics.

Matthias Huck and Hermann Ney. 2012. Pivot Lightly-Supervised Training for Statistical Machine Translation. In *Proceedings of the Tenth Conference of the Association for Machine Translation in the Americas (AMTA 2012)*, San Diego, CA, USA.

Matthias Huck, Simon Riess, and Alexander Fraser. 2017a. Target-side Word Segmentation Strategies for Neural Machine Translation. In *Proceedings of the Second Conference on Machine Translation*, Copenhagen, Denmark. Association for Computational Linguistics.

Matthias Huck, Aleš Tamchyna, Ondřej Bojar, and Alexander Fraser. 2017b. Producing Unseen Morphological Variants in Statistical Machine Translation. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 2, Short Papers*, pages 369–375, Valencia, Spain. Association for Computational Linguistics.

Matthias Huck, David Vilar, Daniel Stein, and Hermann Ney. 2011. Lightly-Supervised Training for Hierarchical Phrase-Based Machine Translation. In *Proceedings of the First workshop on Unsupervised Learning in NLP*, pages 91–96, Edinburgh, Scotland. Association for Computational Linguistics.

Diederik P. Kingma and Jimmy Ba. 2015. Adam: A method for stochastic optimization. *CoRR*, abs/1412.6980.

Philipp Koehn. 2005. Europarl: A Parallel Corpus for Statistical Machine Translation. In *Proceedings of the MT Summit X*, Phuket, Thailand.

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. Association for Computational Linguistics.

Philipp Koehn and Kevin Knight. 2003. Empirical Methods for Compound Splitting. In *Proceedings of the 10th Conference of the European Chapter of the Association for Computational Linguistics*, pages 187–194, Budapest, Hungary. Association for Computational Linguistics.

Patrik Lambert, Holger Schwenk, Christophe Servan, and Sadaf Abdul-Rauf. 2011. Investigations on Translation Model Adaptation Using Monolingual Data. In *Proceedings of the Sixth Workshop on Statistical Machine Translation*, pages 284–293, Edinburgh, Scotland. Association for Computational Linguistics.

Lemao Liu, Masao Utiyama, Andrew Finch, and Eiichiro Sumita. 2016. Agreement on Target-bidirectional Neural Machine Translation. In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 411–416, San Diego, CA, USA. 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, PA, USA. Association for Computational Linguistics.

Jan-Thorsten Peter, Tamer Alkhouli, Hermann Ney, Matthias Huck, Fabienne Braune, Alexander Fraser, Aleš Tamchyna, Ondřej Bojar, Barry Haddow, Rico Sennrich, Frédéric Blain, Lucia Specia, Jan Niehues, Alex Waibel, Alexandre Allauzen, Lauriane Aufrant, Franck Burlot, elena knyazeva, Thomas Lavergne, François Yvon, Mārcis Pinnis, and Stella Frank. 2016. The QT21/HimL Combined Machine Translation System. In *Proceedings of the First Conference on Machine Translation*, pages 344–355, Berlin, Germany. Association for Computational Linguistics.

Anita Ramm and Alexander Fraser. 2016. Modeling verbal inflection for English to German SMT. In *Proceedings of the First Conference on Machine Translation*, pages 21–31, Berlin, Germany. Association for Computational Linguistics.

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, pages 219–224, Sofia, Bulgaria. Association for Computational Linguistics.

Rico Sennrich, Orhan Firat, Kyunghyun Cho, Alexandra Birch, Barry Haddow, Julian Hitschler, Marcin Junczys-Dowmunt, Samuel Läubli, Antonio Valerio Miceli Barone, Jozef Mokry, and Maria Nadejde. 2017. Nematus: a Toolkit for Neural Machine Translation. In *Proceedings of the Software Demonstrations of the 15th Conference of the European Chapter of the Association for Computational Linguistics*, pages 65–68, Valencia, Spain. Association for Computational Linguistics.

Rico Sennrich, Barry Haddow, and Alexandra Birch. 2016a. Edinburgh Neural Machine Translation Systems for WMT 16. In *Proceedings of the First Conference on Machine Translation*, pages 371–376, Berlin, Germany. Association for Computational Linguistics.

Rico Sennrich, Barry Haddow, and Alexandra Birch. 2016b. Improving Neural Machine Translation Models with Monolingual Data. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 86–96, Berlin, Germany. Association for Computational Linguistics.

Rico Sennrich, Barry Haddow, and Alexandra Birch. 2016c. Neural Machine Translation of Rare Words with Subword Units. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1715–1725, Berlin, Germany. Association for Computational Linguistics.

Ilya Sutskever, Oriol Vinyals, and Quoc V Le. 2014. Sequence to Sequence Learning with Neural Networks. In Z. Ghahramani, M. Welling, C. Cortes, N. D. Lawrence, and K. Q. Weinberger, editors, *Advances in Neural Information Processing Systems 27*, pages 3104–3112. Curran Associates, Inc.

Aleš Tamchyna, Alexander Fraser, Ondřej Bojar, and Marcin Junczys-Dowmunt. 2016a. 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. Association for Computational Linguistics.

Aleš Tamchyna, Roman Sudarikov, Ondřej Bojar, and Alexander Fraser. 2016b. CUNI-LMU Submissions in WMT2016: Chimera Constrained and Beaten. In *Proceedings of the First Conference on Machine Translation*, pages 385–390, Berlin, Germany. Association for Computational Linguistics.

Nicola Ueffing, Gholamreza Haffari, and Anoop Sarkar. 2007. Semi-supervised model adaptation for statistical machine translation. *Machine Translation*, 21(2):77–94.

Marion Weller, Max Kisselew, Svetlana Smekalova, Alexander Fraser, Helmut Schmid, Nadir Durrani, Hassan Sajjad, and Richárd Farkas. 2013. Munich-Edinburgh-Stuttgart Submissions at WMT13: Morphological and Syntactic Processing for SMT. In *Proceedings of the Eighth Workshop on Statistical Machine Translation*, pages 232–239, Sofia, Bulgaria. Association for Computational Linguistics.

Marion Weller-Di Marco, Alexander Fraser, and Sabine Schulte im Walde. 2016. Modeling Complement Types in Phrase-Based SMT. In *Proceedings of the First Conference on Machine Translation*, pages 43–53, Berlin, Germany. Association for Computational Linguistics.

Philip Williams, Rico Sennrich, Maria Nadejde, Matthias Huck, Barry Haddow, and Ondřej Bojar. 2016. Edinburgh's Statistical Machine Translation Systems for WMT16. In *Proceedings of the First Conference on Machine Translation*, pages 399–410, Berlin, Germany. Association for Computational Linguistics.

Antonio Jimeno Yepes, Aurélie Névéol, Mariana Neves, Karin Verspoor, Ondřej Bojar, Arthur Boyer, Cristian Grozea, Barry Haddow, Madeleine Kittner, Yvonne Lichtblau, Pavel Pecina, Roland Roller, Amy Siu, Philippe Thomas, and Saskia Trescher. 2017. Findings of the WMT 2017 Biomedical Translation Shared Task. In *Proceedings of the Second Conference on Machine Translation*, Copenhagen, Denmark. Association for Computational Linguistics.