Deepfix: Statistical Post-editing of Statistical Machine Translation Using Deep Syntactic Analysis

Rudolf Rosa and **David Mareček** and **Aleš Tamchyna**
Charles University in Prague, Faculty of Mathematics and Physics
Institute of Formal and Applied Linguistics
Malostranské náměstí 25, Prague
{rosa,marecek,tamchyna}@ufal.mff.cuni.cz

Abstract

Deepfix is a statistical post-editing system for improving the quality of statistical machine translation outputs. It attempts to correct errors in verb-noun valency using deep syntactic analysis and a simple probabilistic model of valency. On the English-to-Czech translation pair, we show that statistical post-editing of statistical machine translation leads to an improvement of the translation quality when helped by deep linguistic knowledge.

1 Introduction

Statistical machine translation (SMT) is the current state-of-the-art approach to machine translation – see e.g. Callison-Burch et al. (2011). However, its outputs are still typically significantly worse than human translations, containing various types of errors (Bojar, 2011b), both in lexical choices and in grammar.

As shown by many researchers, e.g. Bojar (2011a), incorporating deep linguistic knowledge directly into a translation system is often hard to do, and seldom leads to an improvement of translation output quality. It has been shown that it is often easier to correct the machine translation outputs in a second-stage post-processing, which is usually referred to as automatic post-editing.

Several types of errors can be fixed by employing rule-based post-editing (Rosa et al., 2012b), which can be seen as being orthogonal to the statistical methods employed in SMT and thus can capture different linguistic phenomena easily.

But there are still other errors that cannot be corrected with hand-written rules, as there exist many linguistic phenomena that can never be fully described manually – they need to be handled statistically by automatically analyzing large-scale text corpora. However, to the best of our knowledge, there is very little successful research in statistical post-editing (SPE) of SMT (see Section 2).

English	Czech	
go to the doctor	jít k doktorovi	dative case
go to the centre	jít do centra	genitive case
go to a concert	jít na koncert	accusative case
go for a drink	jít na drink	accusative case
go up the hill	jít na kopec	accusative case

Table 1: Examples of valency of the verb 'to go' and 'jít'. For Czech, the morphological cases of the nouns are also indicated.

Source:	The government **spends on** the middle **schools**.
Moses:	Vláda **utrácí** střední **školy**.
Meaning:	The government **destroys** the middle **schools**.
Reference:	Vláda **utrácí za** střední **školy**.
Meaning:	The government **spends on** the middle **schools**.

Table 2: Example of a valency error in output of Moses SMT system.

In our paper, we describe a statistical approach to correcting one particular type of English-to-Czech SMT errors – errors in the verb-noun *valency*. The term *valency* stands for the way in which verbs and their arguments are used together, usually together with prepositions and morphological cases, and is described in Section 4. Several examples of the valency of the English verb 'to go' and the corresponding Czech verb 'jít' are shown in Table 1.

We conducted our experiments using a state-of-the-art SMT system Moses (Koehn et al., 2007). An example of Moses making a valency error is translating the sentence 'The government spends on the middle schools.', adapted from our development data set. As shown in Table 2, Moses translates the sentence incorrectly, making an error in the valency of the 'utrácet – škola' ('spend – school') pair. The missing preposition changes the meaning dramatically, as the verb 'utrácet' is pol-

Proceedings of the ACL Student Research Workshop, pages 172–179,
Sofia, Bulgaria, August 4-9 2013.

ysemous and can mean 'to spend (esp. money)' as well as 'to kill, to destroy (esp. animals)'.

Our approach is to use deep linguistic analysis to automatically determine the structure of each sentence, and to detect and correct valency errors using a simple statistical valency model. We describe our approach in detail in Section 5.

We evaluate and discuss our experiments in Section 6. We then conclude the paper and propose areas to be researched in future in Section 7.

2 Related Work

The first reported results of automatic post-editing of machine translation outputs are (Simard et al., 2007) where the authors successfully performed statistical post-editing (SPE) of rule-based machine translation outputs. To perform the post-editing, they used a phrase-based SMT system in a monolingual setting, trained on the outputs of the rule-based system as the source and the human-provided reference translations as the target, to achieve massive translation quality improvements. The authors also compared the performance of the post-edited rule-based system to directly using the SMT system in a bilingual setting, and reported that the SMT system alone performed worse than the post-edited rule-based system. They then tried to post-edit the bilingual SMT system with another monolingual instance of the same SMT system, but concluded that no improvement in quality was observed.

The first known positive results in SPE of SMT are reported by Oflazer and El-Kahlout (2007) on English to Turkish machine translation. The authors followed a similar approach to Simard et al. (2007), training an SMT system to post-edit its own output. They use two iterations of post-editing to get an improvement of 0.47 BLEU points (Papineni et al., 2002). The authors used a rather small training set and do not discuss the scalability of their approach.

To the best of our knowledge, the best results reported so far for SPE of SMT are by Béchara et al. (2011) on French-to-English translation. The authors start by using a similar approach to Oflazer and El-Kahlout (2007), getting a statistically significant improvement of 0.65 BLEU points. They then further improve the performance of their system by adding information from the source side into the post-editing system by concatenating some of the translated words with their source words, eventually reaching an improvement of 2.29 BLEU points. However, similarly to Oflazer and El-Kahlout (2007), the training data used are very small, and it is not clear how their method scales on larger training data.

Direction	Baseline	SPE	Context SPE
en→cs	**10.85±0.47**	10.70±0.44	10.73±0.49
cs→en	**17.20±0.53**	17.11±0.52	17.18±0.54

Table 3: Results of SPE approach of Béchara et al. (2011) evaluated on English-Czech SMT.

In our previous work (Rosa et al., 2012b), we explored a related but substantially different area of *rule-based* post-editing of SMT. The resulting system, Depfix, manages to significantly improve the quality of several SMT systems outputs, using a set of hand-written rules that detect and correct grammatical errors, such as agreement violations. Depfix can be easily combined with Deepfix,[1] as it is able to correct different types of errors.

3 Evaluation of Existing SPE Approaches

First, we evaluated the utility of the approach of Béchara et al. (2011) for the English-Czech language pair. We used 1 million sentence pairs from CzEng 1.0 (Bojar et al., 2012b), a large English-Czech parallel corpus. Identically to the paper, we split the training data into 10 parts, trained 10 systems (each on nine tenths of the data) and used them to translate the remaining part. The second step was then trained on the concatenation of these translations and the target side of CzEng. We also implemented the *contextual* variant of SPE where words in the intermediate language are annotated with corresponding source words if the alignment strength is greater than a given threshold. We limited ourselves to the threshold value 0.8, for which the best results are reported in the paper. We tuned all systems on the dataset of WMT11 (Callison-Burch et al., 2011) and evaluated on the WMT12 dataset (Callison-Burch et al., 2012).

Table 3 summarizes our results. The reported confidence intervals were estimated using bootstrap resampling (Koehn, 2004). SPE did not lead to any improvements of BLEU in our experiments. In fact, SPE even slightly decreased the score (but

[1] **Dep**fix (Rosa et al., 2012b) performs rule-based post-editing on shallow-syntax **dep**endency trees, while **Deep**fix (described in this paper) is a statistical post-editing system operating on **deep**-syntax dependency trees.

the difference is statistically insignificant in all cases).

We conclude that this method does not improve English-Czech translation, possibly because our training data is too large for this method to bring any benefit. We therefore proceed with a more complex approach which relies on deep linguistic knowledge.

4 Deep Dependency Syntax, Formemes, and Valency

4.1 Tectogrammatical dependency trees

Tectogrammatical trees are deep syntactic dependency trees based on the Functional Generative Description (Sgall et al., 1986). Each node in a tectogrammatical tree corresponds to a content word, such as a noun, a full verb or an adjective; the node consists of the lemma of the content word and several other attributes. Functional words, such as prepositions or auxiliary verbs, are not directly present in the tectogrammatical tree, but are represented by attributes of the respective content nodes. See Figure 1 for an example of two tectogrammatical trees (for simplicity, most of the attributes are not shown).

In our work, we only use one of the many attributes of tectogrammatical nodes, called *formeme* (Dušek et al., 2012). A formeme is a string representation of selected morpho-syntactic features of the content word and selected auxiliary words that belong to the content word, devised to be used as a simple and efficient representation of the node.

A noun formeme, which we are most interested in, consists of three parts (examples taken from Figure 1):

1. The syntactic part-of-speech – **`n`** for nouns.
2. The preposition if the noun has one (empty otherwise), as in `n:`**`on`**`+X` or `n:`**`za`**`+4`.
3. A form specifier.
   - In English, it typically marks the subject or object, as in `n:`**`subj`**. In case of a noun accompanied by a preposition, the third part is always `X`, as in `n:on+`**`X`**.
   - In Czech, it denotes the morphological case of the noun, represented by its number (from 1 to 7 as there are seven cases in Czech), as in `n:`**`1`** and `n:za+`**`4`**.

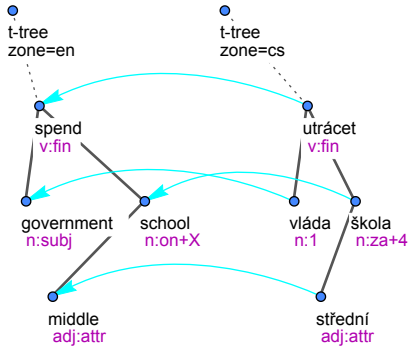

Figure 1: Tectogrammatical trees for the sentence 'The government spends on the middle schools.' – 'Vláda utrácí za střední školy.'; only lemmas and formemes of the nodes are shown.

Adjectives and nouns can also have the `adj:attr` and `n:attr` formemes, respectively, meaning that the node is in morphological agreement with its parent. This is especially important in Czech, where this means that the word bears the same morphological case as its parent node.

4.2 Valency

The notion of *valency* (Tesnière and Fourquet, 1959) is semantic, but it is closely linked to syntax. In the theory of valency, each verb has one or more *valency frames*. Each valency frame describes a meaning of the verb, together with arguments (usually nouns) that the verb must or can have, and each of the arguments has one or several fixed forms in which it must appear. These forms can typically be specified by prepositions and morphological cases to be used with the noun, and thus can be easily expressed by formemes.

For example, the verb 'to go', shown in Table 1, has a valency frame that can be expressed as `n:subj go n:to+X`, meaning that the subject goes to some place.

The valency frames of the verbs 'spend' and 'utrácet' in Figure 1 can be written as `n:subj spend n:on+X` and `n:1 utrácet n:za+4`; the subject (in Czech this is a noun in nominative case) spends on an object (in Czech, the preposition 'za' plus a noun in accusative case).

In our work, we have extended our scope also to noun-noun valency, i.e. the parent node can be either a verb or a noun, while the arguments are always nouns. Practice has proven this extension to be useful, although the majority of the corrections

performed are still of the verb-noun valency type. Still, we keep the traditional notion of verb-noun valency throughout the text, especially to be able to always refer to the parent as "the verb" and to the child as "the noun".

5 Our Approach

5.1 Valency models

To be able to detect and correct valency errors, we created statistical models of verb-noun valency. We model the conditional probability of the noun argument formeme based on several features of the verb-noun pair. We decided to use the following two models:

$$P(f_n|l_v, f_{EN}) \quad (1)$$

$$P(f_n|l_v, l_n, f_{EN}) \quad (2)$$

where:

- f_n is the formeme of the Czech noun argument, which is being modelled
- l_v is the lemma of the Czech parent verb
- l_n is the lemma of the Czech noun argument
- f_{EN} is the formeme of the English noun aligned to the Czech noun argument

The input is first processed by the model (1), which performs more general fixes, in situations where the (l_v, f_{EN}) pair rather unambiguously defines the valency frame required.

Then model (2) is applied, correcting some errors of the model (1), in cases where the noun argument requires a different valency frame than is usual for the (l_v, f_{EN}) pair, and making some more fixes in cases where the correct valency frame required for the (l_v, f_{EN}) pair was too ambiguous to make a correction according to model (1), but the decision can be made once information about l_n is added.

We computed the models on the full training set of CzEng 1.0 (Bojar et al., 2012b) (roughly 15 million sentences), and smoothed the estimated probabilities with add-one smoothing.

5.2 Deepfix

We introduce a new statistical post-editing system, Deepfix, whose input is a pair of an English sentence and its Czech machine translation, and the output is the Czech sentence with verb-noun valency errors corrected.

The Deepfix pipeline consists of several steps:

1. the sentences are tokenized, tagged and lemmatized (a lemma and a morphological tag is assigned to each word)
2. corresponding English and Czech words are aligned based on their lemmas
3. deep-syntax dependency parse trees of the sentences are built, the nodes in the trees are labelled with formemes
4. improbable noun formemes are replaced with correct formemes according to the valency model
5. the words are regenerated according to the new formemes
6. the regenerating continues recursively to children of regenerated nodes if they are in morphological agreement with their parents (which is typical for adjectives)

To decide whether the formeme of the noun is incorrect, we query the valency model for all possible formemes and their probabilities. If an alternative formeme probability exceeds a fixed threshold, we assume that the original formeme is incorrect, and we use the alternative formeme instead.

For our example sentence, 'The government spends on the middle schools.' – 'Vláda utrácí za střední školy.', we query the model (2) and get the following probabilities:

- P(n:4 | utrácet, škola, n:on+X) = 0.07
  (the original formeme)
- P(n:za+4 | utrácet, škola, n:on+X) = 0.89
  (the most probable formeme)

The threshold for this change type is 0.86, is exceeded by the `n:za+4` formeme and thus the change is performed: 'školy' is replaced by 'za školy'.

5.3 Tuning the Thresholds

We set the thresholds differently for different types of changes. The values of the thresholds that we used are listed in Table 4 and were estimated manually. We distinguish changes where only the morphological case of the noun is changed from changes to the preposition. There are three possible types of a change to a preposition: switching one preposition to another, adding a new preposition, and removing an existing preposition. The

Correction type	Thresholds for models (1)	(2)
Changing the noun case only	0.55	0.78
Changing the preposition	0.90	0.84
Adding a new preposition	–	0.86
Removing the preposition	–	–

Table 4: Deepfix thresholds

change to the preposition can also involve changing the morphological case of the noun, as each preposition typically requires a certain morphological case.

For some combinations of a change type and a model, as in case of the preposition removing, we never perform a fix because we observed that it nearly never improves the translation. E.g., if a verb-noun pair can be correct both with and without a preposition, the preposition-less variant is usually much more frequent than the prepositional variant (and thus is assigned a much higher probability by the model). However, the preposition often bears a meaning that is lost by removing it – in Czech, which is a relatively free-word-order language, the semantic roles of verb arguments are typically distinguished by prepositions, as opposed to English, where they can be determined by their relative position to the verb.

5.4 Implementation

The whole Deepfix pipeline is implemented in Treex, a modular NLP framework (Popel and Žabokrtský, 2010) written in Perl, which provides wrappers for many state-of-the-art NLP tools. For the analysis of the English sentence, we use the Morče tagger (Spoustová et al., 2007) and the MST parser (McDonald et al., 2005). The Czech sentence is analyzed by the Featurama tagger[2] and the RUR parser (Rosa et al., 2012a) – a parser adapted to parsing of SMT outputs. The word alignment is created by GIZA++ (Och and Ney, 2003); the intersection symmetrization is used.

6 Evaluation

6.1 Automatic Evaluation

We evaluated our method on three datasets: WMT10 (2489 parallel sentences), WMT11 (3003 parallel sentences), and WMT12 (3003 parallel sentences) by Callison-Burch et al. (2010; 2011; 2012). For evaluation, we used outputs of a state-of-the-art SMT system, Moses (Koehn et al., 2007), tuned for English-to-Czech translation (Bojar et al., 2012a). We used the WMT10 dataset and its Moses translation as our development data to tune the thresholds. In Table 5, we report the achieved BLEU scores (Papineni et al., 2002), NIST scores (Doddington, 2002), and PER (Tillmann et al., 1997).

The improvements in automatic scores are low but consistently positive, which suggests that Deepfix does improve the translation quality. However, the changes performed by Deepfix are so small that automatic evaluation is unable to reliably assess whether they are positive or negative – it can only be taken as an indication.

6.2 Manual Evaluation

To reliably assess the performance of Deepfix, we performed manual evaluation on the WMT12 dataset translated by the Moses system.

The dataset was evenly split into 4 parts and each of the parts was evaluated by one of two annotators (denoted "A" and "B"). For each sentence that was modified by Deepfix, the annotator decided whether the Deepfix correction had a positive ("improvement") or negative ("degradation") effect on the translation quality, or concluded that this cannot be decided ("indefinite") – either because both of the sentences are correct variants, or because both are incorrect.[3]

The results in Table 6 prove that the overall effect of Deepfix is positive: it modifies about 20% of the sentence translations (569 out of 3003 sentences), improving over a half of them while leading to a degradation in only a quarter of the cases.

We measured the inter-annotator agreement on 100 sentences which were annotated by both annotators. For 60 sentence pairs, both of the annotators were able to select which sentence is better, i.e. none of the annotators used the "indefinite" marker. The inter-annotator agreement on these 60 sentence pairs was 97%.[4]

[2] http://featurama.sourceforge.net/

[3] The evaluation was done in a blind way, i.e. the annotators did not know which sentence is before Deepfix and which is after Deepfix. They were also provided with the source English sentences and the reference human translations.

[4] If all 100 sentence pairs are taken into account, requiring that the annotators also agree on the "indefinite" marker, the inter-annotator agreement is only 65%. This suggests that deciding whether the translation quality differs significantly is much harder than deciding which translation is of a higher quality.

Dataset	BLEU score (higher is better)			NIST score (higher is better)			PER (lower is better)		
	Baseline	Deepfix	Difference	Baseline	Deepfix	Difference	Baseline	Deepfix	Difference
WMT10*	15.66	15.74	+0.08	5.442	5.470	+0.028	58.44%	58.26%	-0.18
WMT11	16.39	16.42	+0.03	5.726	5.737	+0.011	57.17%	57.09%	-0.08
WMT12	13.81	13.85	+0.04	5.263	5.283	+0.020	60.04%	59.91%	-0.13

Table 5: Automatic evaluation of Deepfix on outputs of the Moses system on WMT10, WMT11 and WMT12 datasets. *Please note that WMT10 was used as the development dataset.

Part	Annotator	Changed sentences	Improvement	Degradation	Indefinite
1	A	126	57 (45%)	35 (28%)	34 (27%)
2	B	112	62 (55%)	29 (26%)	21 (19%)
3	A	150	88 (59%)	29 (19%)	33 (22%)
4	B	181	114 (63%)	42 (23%)	25 (14%)
Total		569	321 (56%)	135 (24%)	113 (20%)

Table 6: Manual evaluation of Deepfix on outputs of Moses Translate system on WMT12 dataset.

6.3 Discussion

When a formeme change was performed, it was usually either positive or at least not harmful (substituting one correct variant for another correct variant).

However, we also observed a substantial amount of cases where the change of the formeme was incorrect. Manual inspection of a sample of these cases showed that there can be several reasons for a formeme change to be incorrect:

- incorrect analysis of the Czech sentence
- incorrect analysis of the English sentence
- the original formeme is a correct but very rare variant

The most frequent issue is the first one. This is to be expected, as the Czech sentence is often erroneous, whereas the NLP tools that we used are trained on correct sentences; in many cases, it is not even clear what a correct analysis of an incorrect sentence should be.

7 Conclusion and Future Work

On the English-Czech pair, we have shown that statistical post-editing of statistical machine translation outputs is possible, even when translating from a morphologically poor to a morphologically rich language, if it is grounded by deep linguistic knowledge. With our tool, Deepfix, we have achieved improvements on outputs of two state-of-the-art SMT systems by correcting verb-noun valency errors, using two simple probabilistic valency models computed on large-scale data. The improvements have been confirmed by manual evaluation.

We encountered many cases where the performance of Deepfix was hindered by errors of the underlying tools, especially the taggers, the parsers and the aligner. Because the use of the RUR parser (Rosa et al., 2012a), which is partially adapted to SMT outputs parsing, lead to a reduction of the number of parser errors, we find the approach of adapting the tools for this specific kind of data to be promising.

We believe that our method can be adapted to other language pairs, provided that there is a pipeline that can analyze at least the target language up to deep syntactic trees. Because we only use a small subset of information that a tectogrammatical tree provides, it is sufficient to use only simplified tectogrammatical trees. These could be created by a small set of rules from shallow-syntax dependency trees, which can be obtained for many languages using already existing parsers.

Acknowledgments

This research has been supported by the 7th FP project of the EC No. 257528 and the project 7E11042 of the Ministry of Education, Youth and Sports of the Czech Republic.

Data and some tools used as a prerequisite for the research described herein have been provided by the LINDAT/CLARIN Large Infrastructural project, No. LM2010013 of the Ministry of Education, Youth and Sports of the Czech Republic.

We would like to thank two anonymous reviewers for many useful comments on the manuscript of this paper.

References

Hanna Béchara, Yanjun Ma, and Josef van Genabith. 2011. Statistical post-editing for a statistical MT system. *MT Summit XIII*, pages 308–315.

Ondřej Bojar, Bushra Jawaid, and Amir Kamran. 2012a. Probes in a taxonomy of factored phrase-based models. In *Proceedings of the Seventh Workshop on Statistical Machine Translation*, pages 253–260, Montréal, Canada. Association for Computational Linguistics.

Ondřej Bojar, Zdeněk Žabokrtský, Ondřej Dušek, Petra Galuščáková, Martin Majliš, David Mareček, Jiří Maršík, Michal Novák, Martin Popel, and Aleš Tamchyna. 2012b. The joy of parallelism with CzEng 1.0. In *Proceedings of the 8th International Conference on Language Resources and Evaluation (LREC 2012)*, pages 3921–3928, İstanbul, Turkey. European Language Resources Association.

Ondřej Bojar. 2011a. Rich morphology and what can we expect from hybrid approaches to MT. Invited talk at International Workshop on Using Linguistic Information for Hybrid Machine Translation (LIHMT-2011), November.

Ondřej Bojar. 2011b. Analyzing error types in English-Czech machine translation. *Prague Bulletin of Mathematical Linguistics*, 95:63–76, March.

Chris Callison-Burch, Philipp Koehn, Christof Monz, Kay Peterson, Mark Przybocki, and Omar Zaidan. 2010. Findings of the 2010 joint workshop on statistical machine translation and metrics for machine translation. In *Proceedings of the Joint Fifth Workshop on Statistical Machine Translation and MetricsMATR*, pages 17–53, Uppsala, Sweden, July. Association for Computational Linguistics.

Chris Callison-Burch, Philipp Koehn, Christof Monz, and Omar Zaidan. 2011. Findings of the 2011 workshop on statistical machine translation. In *Proceedings of the Sixth Workshop on Statistical Machine Translation*, pages 22–64, Edinburgh, Scotland, July. Association for Computational Linguistics.

Chris Callison-Burch, Philipp Koehn, Christof Monz, Matt Post, Radu Soricut, and Lucia Specia. 2012. Findings of the 2012 workshop on statistical machine translation. In *Proceedings of the Seventh Workshop on Statistical Machine Translation*, pages 10–51, Montréal, Canada, June. Association for Computational Linguistics.

George Doddington. 2002. Automatic evaluation of machine translation quality using n-gram co-occurrence statistics. In *Proceedings of the second international conference on Human Language Technology Research*, pages 138–145. Morgan Kaufmann Publishers Inc.

Ondřej Dušek, Zdeněk Žabokrtský, Martin Popel, Martin Majliš, Michal Novák, and David Mareček. 2012. Formemes in English-Czech deep syntactic MT. In *Proceedings of the Seventh Workshop on Statistical Machine Translation*, pages 267–274, Montréal, Canada. 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, Ondřej Bojar, Alexandra Constantin, and Evan Herbst. 2007. Moses: Open source toolkit for statistical machine translation. In *ACL 2007, 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.

Philipp Koehn. 2004. Statistical significance tests for machine translation evaluation. In *Proc. of EMNLP*, Barcelona, Spain.

Ryan McDonald, Koby Crammer, and Fernando Pereira. 2005. Online large-margin training of dependency parsers. In *Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics*, pages 91–98. Association for Computational Linguistics.

Franz Josef Och and Hermann Ney. 2003. A systematic comparison of various statistical alignment models. *Computational Linguistics*, 29(1):19–51.

Kemal Oflazer and Ilknur Durgar El-Kahlout. 2007. Exploring different representational units in English-to-Turkish statistical machine translation. In *Proceedings of the Second Workshop on Statistical Machine Translation*, pages 25–32. 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 *ACL 2002, Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics*, pages 311–318, Philadelphia, Pennsylvania.

Martin Popel and Zdeněk Žabokrtský. 2010. TectoMT: modular NLP framework. In *Proceedings of the 7th international conference on Advances in natural language processing*, IceTAL'10, pages 293–304, Berlin, Heidelberg. Springer-Verlag.

Rudolf Rosa, Ondřej Dušek, David Mareček, and Martin Popel. 2012a. Using parallel features in parsing of machine-translated sentences for correction of grammatical errors. In *Proceedings of Sixth Workshop on Syntax, Semantics and Structure in Statistical Translation (SSST-6), ACL*, pages 39–48, Jeju, Korea. ACL.

Rudolf Rosa, David Mareček, and Ondřej Dušek. 2012b. DEPFIX: A system for automatic correction of Czech MT outputs. In *Proceedings of the Seventh Workshop on Statistical Machine Translation*,

pages 362–368, Montréal, Canada. Association for Computational Linguistics.

Petr Sgall, Eva Hajičová, and Jarmila Panevová. 1986. *The meaning of the sentence in its semantic and pragmatic aspects*. Springer.

Michel Simard, Cyril Goutte, and Pierre Isabelle. 2007. Statistical phrase-based post-editing. In *Human Language Technologies 2007: The Conference of the North American Chapter of the Association for Computational Linguistics; Proceedings of the Main Conference*, pages 508–515, Rochester, New York, April. Association for Computational Linguistics.

Drahomíra Spoustová, Jan Hajič, Jan Votrubec, Pavel Krbec, and Pavel Květoň. 2007. The best of two worlds: Cooperation of statistical and rule-based taggers for Czech. In *Proceedings of the Workshop on Balto-Slavonic Natural Language Processing 2007*, pages 67–74, Praha, Czechia. Univerzita Karlova v Praze, Association for Computational Linguistics.

Lucien Tesnière and Jean Fourquet. 1959. *Eléments de syntaxe structurale*. Éditions Klincksieck, Paris.

Christoph Tillmann, Stephan Vogel, Hermann Ney, Alex Zubiaga, and Hassan Sawaf. 1997. Accelerated dp based search for statistical translation. In *European Conf. on Speech Communication and Technology*, pages 2667–2670.