A Comparison of MT Methods for Closely Related Languages: a Case Study on Czech – Slovak Language Pair [*]

Vladislav Kuboň[1]
[1]Institute of Formal and Applied Linguistics
Faculty of Mathematics and Physics
Charles University in Prague
vk@ufal.mff.cuni.cz

Jernej Vičič[2,3]
[2]FAMNIT
University of Primorska,
[3]Fran Ramovš Institute of the
Slovenian Language, SRC SASA,
jernej.vicic@upr.si

Abstract

This paper describes an experiment comparing results of machine translation between two closely related languages, Czech and Slovak. The comparison is performed by means of two MT systems, one representing rule-based approach, the other one representing statistical approach to the task. Both sets of results are manually evaluated by native speakers of the target language. The results are discussed both from the linguistic and quantitative points of view.

1 Introduction

Machine translation (MT) of related languages is a specific field in the domain of MT which attracted the attention of several research teams in the past by promising relatively good results through the application of classic rule-based methods. The exploitation of lexical, morphological and syntactic similarity of related languages seemed to balance the advantages of data-driven approaches, especially for the language pairs with smaller volumes of available parallel data.

This simple and straightforward assumption have led to the construction of numerous rule-based translation systems for related (or similar) natural languages. The following list (ordered alphabetically) includes several examples of those systems:

- (Altintas and Cicekli, 2002) for Turkic languages.

- Apertium (Corbi-Bellot et al., 2005) for Romance languages.

- (Dyvik, 1995; Bick and Nygaard, 2007; Ahrenberg and Holmqvist, 2004) for Scandinavian languages.

- Česílko (Hajič et al., 2000), for Slavic languages with rich inflectional morphology, mostly language pairs with Czech language as a source.

- Ruslan (Oliva, 1989) full-fledged transfer based RBMT system from Czech to Russian.

- (Scannell, 2006) for Gaelic languages; Irish (Gaeilge) and Scottish Gaelic (G'aidhlig).

- (Tyers et al., 2009) for the North Sami to Lule Sami language pair.

- Guat (Vičič, 2008) for Slavic languages with rich inflectional morphology, mostly language pairs with Slovenian language.

Many of the systems listed above had been created in the period when it was hard to obtain a good quality data-driven system which would enable comparison against these systems. The existence of Google Translate[1] which nowadays enables the automatic translation even between relatively small languages made it possible to investigate advantages and disadvantages of both approaches. This paper introduces the first step in this direction - the comparison of results of two different systems for two really very closely related languages - Czech and Slovak.

2 State of the art

There has already been a lot of research in Machine Translation evaluation. There are quite a few conferences and shared tasks devoted entirely to this problem such as NIST Machine Translation Evaluation (NIST, 2009) or Workshop on Statistical Machine Translation (Bojar et al., 2013). (Weijnitz et al., 2004) presents a research on how systems from two different MT paradigms cope with a new domain. (Kolovratník et al., 2009) presents a research on how relatedness of languages influences the translation quality of a SMT sytem.

The novelty of the presented paper is in the focus on machine translation for closely related languages and in the comparison of the two mostly used paradigms for this task: shallow parse and transfer RBMT and SMT paradigms.

3 Translation systems

The translation systems selected for the experiment are:

- Google Translate

[1]Google Translate: https://translate.google.com/.

- Česílko (Hajič et al., 2003)

Google Translate was selected as the most used translation system. Česílko belongs to the shallow-parse and shallow-transfer rule based machine translation paradigm which is by many authors the most suitable for translation of related languages.

Česílko (Hajič et al., 2003) was used as a representative of rule-based MT systems, the translation direction from Czech to Slovak was naturally chosen because this is the only direction this system supports for this particular language pair.

The on-line publicly available versions of the systems sere used in the experiment to ensure the reproducibility of the experiment. All the test data is publicly available at the language technologies server of the University of Primorska[2].

Let us now introduce the systems in a more detail.

3.1 Google Translate

This system is currently probably the most popular and most widely used MT system in the world. It belongs to the Statistical Machine Translation – SMT paradigm. SMT is based on parametric statistical models, which are constructed on bilingual aligned corpora (training data). The methods focus on looking for general patterns that arise in the use of language instead of analyzing sentences according to grammatical rules. The main tool for finding such patterns is counting a variety of objects – statistics. The main idea of the paradigm is to model the probability that parts of a sentence from the source language translate into suitable parts of sentence in the target language.

The system takes advantage of the vast parallel resources which Google Inc. has at their disposal and it is therefore able to translate a large number of language pairs. Currently (July 2014), this system offers automatic translation among 80 languages. This makes it a natural candidate as a universal quality standard for MT, especially for pairs of smaller (underrepresented) languages for which there are very few MT systems.

3.2 Česílko

One of the first systems which fully relied on the similarity of related languages, Česílko (Hajič et al., 2003), had originally a very simple architecture. Its first implementation translated from Czech to Slovak. It used the method of direct word-for-word translation (after necessary morphological processing). More precisely, it translated each lemma obtained by morphological analysis and morphological tag provided by a tagger to a lemma and a corresponding tag in the target language. For the translation of lemmas it was necessary to use a bilingual dictionary, the differences in morphology of both languages, although to a large extent regular, did not allow to use a simple transliteration. The translation

of lemmas was necessary due to differences in tagsets of the source and target language.

The syntactic similarity of both languages allowed the omission of syntactic analysis of the source language and syntactic synthesis of the target one, therefore the dictionary phase of the system had been immediately followed by morphological synthesis of the target language. No changes of the word order were necessary, the target language order of words preserved the word order of the source language.

Later versions of the system experimented with the architecture change involving the omission of the source language tagger and the addition of a stochastic ranker of translation hypothesis at the target language side. The purpose of this experiment was to eliminate tagging errors at the beginning of the translation process and to enable more variants of translation from which the stochastic ranker chose the most probable hypothesis. This change has been described for example in (Homola and Kuboň, 2008). For the purpose of our experiment we are using the original version of the system which has undergone some minor improvements (better tagger, improved dictionary etc.) and which is publicly available for testing at the website of the LINDAT project[3]. The decision to use this version is natural, given the fact that this is the only publicly available version of the system.

4 Methodology

In the planning phase of our experiment it was necessary to make a couple of decisions which could cause certain bias and invalidate the results obtained. First of all, the choice of the language pair (Czech to Slovak) was quite natural. These languages show very high degree of similarity at all levels (morphological, syntactic, semantic) and thus they constitute an ideal language pair for the development of simplified rule-based architecture. We are of course aware that for a complete answer to this question it would be necessary to test more systems and more language pairs, but in this phase of our experiments we do not aim at obtaining a complete answer, our main goal is to develop a methodology and to perform some kind of pilot testing showing the possible directions of future research.

The second important decision concerned the method of evaluation. Our primary goal was to set up a method which would be relatively simple and fast, thus allowing to manually (the reasons for manual evaluation are given in 4.2.1 subsection) process reasonable volume of results. The second goal concerned the endeavor to estimate evaluator's confidence in their judgments.

4.1 Basic properties of the language pair

The language pair used in our experiment belongs to western Slavic language group. We must admit that

---

[2] Test data: `http://jt.upr.si/research_projects/related_languages/`

[3] Česílko: `http://lindat.mff.cuni.cz/services/cesilko/`

the reason for choosing this language group was purely pragmatic – there is an extensive previous experience with the translation of several language pairs from this group, see, e.g. (Hajič et al., 2003), (Homola and Kuboň, 2008) or (Homola and Vičič, 2010). On top of that, the availability of the Česílko demo in the LINDAT repository for the free translation of up to 5000 characters naturally led to the decision to use this system (although it is in fact the original version of the system with very simple architecture).

Czech and Slovak represent the closest language pair among the western Slavic languages. Their morphology and syntax are very similar, their lexicons slightly differ, their word order is free (and also similar). In the former Czechoslovakia it was quite common that people understood both languages very well, but after the split of the country the younger people who don't have regular contact with the other language experience certain difficulties because the number of the words unknown to them is not negligible.

However, the greatest challenge for the word-for-word translation approach is not the lexicon (the differences in the lexicon can be handled by a bilingual dictionary), but the ambiguity of word forms. These are typically not part-of-speech ambiguities, they are quite rare although they do exist (*stát* [to stay/the state], *žena* [woman/chasing] or *tři* [three/rub(imper.)]), however, the greatest challenge is the ambiguity of gender, number and case (for example, the form of the adjective *jarní* [spring] is 27-way ambiguous). Resolving this ambiguity is very important for translation because Czech has very strict requirements on agreement (not only subject - predicate agreement, but also agreement in number, gender and case in nominal groups). Even though several Slavic languages including Slovak exhibit similar richness of word forms, the morphological ambiguity is not preserved at all or it is preserved only partially, it is distributed in a different manner and the "form-for-form" translation is not applicable.

For example, if we want to translate the Czech expression *jarní louka* [a spring meadow] into Slovak word for word, it is necessary to disambiguate the adjective which has the same form in Czech for all four genders (in Czech, there are two masculine genders - animate and inanimate) while in Slovak, there are three different forms for masculin, feminin and neutral gender - *jarný, jarná, jarné*. The disambiguation is performed by a state-of-the art stochastic tagger. Although this brings a stochastic factor into the system, we still consider Česílko to be primarily rule based system.

4.2 Experiment outline

The aim of the experiment was double: to show the quality of the simple RBMT methods (shallow-parse and shallow transfer RBMT) in comparison to the state-of-the-art SMT system. The second part of the experiment was to outline the most obvious and most challenging errors produced by each translation paradigm.

4.2.1 Translation quality evaluation

This part of the experiment relied on the methodology similar to that used in the 2013 Workshop on Statistical Machine Translation (Bojar et al., 2013). We conducted manual evaluation of both systems' outputs consisting of ranking individual translated sentences according to the translation quality (the evaluators had access to the original sentence). Unlike the ranking of the SMT Workshop which worked always with 5 translations, our task was much simpler and the ranking naturally consisted of ranking translated sentences of both systems. The evaluator indicated which of the two systems is better, having also the chance to indicate that both translations are identical, because the systems produced relatively large number of identical results - see section 5).

The reason why we didn't automatic measures of translation quality was quite natural. After a period of wide acceptance of automatic measures like BLEU (Papineni et al., 2001) or NIST (NIST, 2009), recent MT evaluation experiments seem to prefer manual methods. Many papers such as Callison-Burch et al. (2006) and authors of workshops such as WMT 2013 (Bojar et al., 2013) contend that automatic measures of machine translation quality are an imperfect substitute for human assessments, especially when it is necessary to compare different systems (or, even worse, the systems based on different paradigms).

4.2.2 Test data

Our evaluation is based upon a small, yet relevant, test corpus. Because one of the systems undergoing the evaluation has been developed by Google, the creation of the test set required special attention. We could not use any already existing on-line corpus as Google regularly enhances language models with new language data. Any on-line available corpus could have already been included in the training data of Google Translate, thus the results of the evaluation would have been biased towards the SMT system. Therefore we have decided to use fresh newspaper texts which cannot be part of any training data set used by Google.

We have selected 200 sentences from fresh newspaper articles of the biggest Czech on-line daily newspapers. Several headline news were selected in order to avoid author bias although the the domain remained daily news. We have selected articles from "iDnes"[4], "Lidovky"[5] and "Novinky"[6]. The test set was created from randomly selected articles on the dates between *14.7.2014* and *18.7.2014*.

All the test-data is publicly available at the language technologies server of the University of Primorska[2].

This part of the experiment consisted in manually examining the translated data from the translation quality evaluation task (described in section 4.2.1). As we have

---

[4]iDnes: `http://www.idnes.cz/`
[5]Lidovky: `http://www.lidovky.cz/`
[6]Novinky: `http://www.novinky.cz/`

expected, the most common errors of Česílko were out of the vocabulary errors. The dictionary coverage of the system has apparently been inadequate for a wide variety of topics from daily news. The results are presented in section 5.1.

5 Results

The results of our experiment are summarized in Table 1. The evaluation has been performed by 5 native speakers of Slovak, the sentences have been randomized so that no evaluator could know which of the two systems produced which translation. The evaluators were asked to mark which translation they consider to be better. Ties were not allowed, but the evaluators were also asked to mark identical sentences. This requirement served also as a kind of thoroughness check, too many unrecognized identical sentences could indicate that the evaluator lost concentration during the task.

	Sent. count	Percentage
Identical sentences	43	21.5%
Clear win of RBMT	10	5%
Clear win of SMT	59	29.5%
Win by voting - RBMT	23	11.5%
Win by voting - SMT	62	31%
Draw	3	1.5%
Total	200	100%

Table 1: Evaluation of results

The rows of Table 1 marked as *Clear win* of one of the systems represent the sentences where none of the evaluators marked the other system as the *better one*. Win by voting does not distinguish how many evaluators were against the system marked by the majority as being the better of the two. The 3 sentences in the Draw row represent the cases when 1 or 3 evaluators mistakenly marked the pair of translations as being identical and there was no majority among the remaining ones.

The results clearly indicate that the quality of Google Translate is better, although it clearly dominates in less than one third of translations. The large number of identical sentences also means that although Česílko produced only 5% of translations which were clearly better than those of Google, it reached absolutely identical quality of translation in yet another 21.5%. This actually means that the top quality translations have been achieved in 26.5% by Česílko and in 51% by Google Translate. According to our opinion, this ratio (approximately 2:1 in favor of the SMT approach) more realistically describes the difference in quality than the ratio of clear wins (approx. 6:1 for Google Translate).

5.1 Errors

This section presents the most obvious errors detected in the evaluation of both systems.

First of all, before we'll look at individual types of errors of both systems, it is necessary to mention one very surprising fact concerning the translations. Although we have expected substantial differences between the corresponding sentences, the translations produced by both systems are surprisingly similar, 21.5% of them being absolutely identical. On top of that, when we have compared the first 100 translated sentences, we have discovered that the edit distance between the two sets is only 493 elementary operations. Given that the translations produced by Google Translate contain 9.653 characters in the first 100 sentences of the test set, this actually represents only about 5% difference.

This looks much more like the results of two variants of the same system than the results of two different systems based upon two completely different paradigms. Because no details about the Google translate for this language pair have been published, it is impossible to judge the reasons for such a similarity. The following example demonstrates this similarity, it represents quite typical example of a long sentence with very few differences between both translations. Errors in translations are stressed by a bold font.

Example 1.

Source: *"V momentě, kdy by třeba Praha chtěla převést systém na Plzeňskou kartu či kartu Českých drah, musela by porušit autorský zákon, protože jedině autor může se softwarem nakládat," vysvětluje mluvčí EMS.*

Google: "V momente, kedy by **potrebné** Praha chcela previesť systém na Plzeňskú kartu či kartu Českých dráh, musela by porušiť autorský zákon, pretože jedine autor môže so softvérom **zaobchádzať**," vysvetľuje hovorca EMS.

Česílko: "V momente, kedy by napríklad Praha chcela previesť systém na Plzenskú **karta** či **karta Český** dráh, musela by porušiť autorský zákon, pretože **jedině** autor môže so softwarom **nakladať**," vysvetľuje hovorca EMS.

English translation: *"In the moment when Prague would like for example to transfer the system to Pilsen card or the Czech railways card, it would have to violate the copyright law, because the author is the only person which can modify the software," explains the speaker of the EMS.*

Even more suspicious are translated sentences which are identical, incorrect and both contain the same error. If something like that happens in a school, the teacher has all reasons to think that one of the two pupils is cheating and that he copied from his neighbor. The systems had no chance to cheat, what makes identical results as in the following example very weird. It would be very interesting to perform more detailed tests in the future and to investigate the reasons for such behavior of two completely different systems. The straightforward explanation that both languages are so similar that

these identical errors simply happen, seems to be too simplistic. Example 2 clearly shows that both systems misinterpreted the Czech adjective *právní* (legal) in the context which allowed reading the first three words of the source sentence as "It is legal" without the regard to the context of the rest of the sentence.

Example 2.

Source: *Je to právní, ale i technologický problém.*

Both systems: Je to **právne**, ale aj technologický problém.

English: *It is a legal, but also a technological problem.*

Let us now look at individual categories of errors.

Lexical errors

The most frequent lexical errors are untranslated words. This happens solely in the translations performed by the RBMT system Česílko due to inadequate coverage of the wide domain of newspaper articles. Some of the cases of untranslated words may have escaped the evaluators̓ attention simply because Česílko leaves out-of-the-vocabulary words unchanged. Because Czech and Slovak are really very close also at the lexical level, some of the word forms used in both languages are identical, and thus they fit into the target sentence. Increasing the coverage of the bilingual dictionary (it currently contains about 40,000 lemmas) would definitely improve the translation quality.

Another lexical error produced entirely by the RBMT system is a wrong translation of some irregular words, as in the following example.

Example 3.

Source: *Mnozí lidé si téměř neumějí představit, že by zapomněli svůj rodný jazyk.*

Google: Mnohí **ľudia** si takmer nevedia predstaviť, že by zabudli svoj rodný jazyk.

Česílko: Mnohí **člověk** si takmer nevedia představit, že by zabudli svoj rodný jazyk.

English translation: *Many people cannot imagine that they could forget their native language.*

The plural of *člověk* [human] is irregular in Czech (*lidé* [people]). Although this error looks like a lexical error, it is more likely caused by the conceptual differences between the morphological analysis of Czech (which recognizes the form as a plural of the lemma *člověk*) and the synthesis of Slovak which uses two lemmas instead of one, one for singular (*člověk*) and one for plural (*ľudia*). The Czech plural word form is then never correctly translated to the Slovak plural form.

Much more serious errors are mistranslated words produced by Google Translate. Such errors are quite typical for phrase-based SMT systems. Let us present an example which appeared a couple of times in our test corpus.

Example 4.

Source: *Česká veřejnost si na omezení zvykla a většinou je respektuje.*

Google: **Slovenská** verejnosť si na obmedzenie zvykla a väčšinou je rešpektuje.

Česílko: **Český** verejnosť si na obmedzenie zvykla a väčšinou ich rešpektuje.

English translation: *Czech public got used to the limits and mostly accepts them.*

The incorrect translation of the adjective *Česká* [Czech] as *Slovenská* [Slovak] has most probably been caused by the language model based upon target language text where the occurrences of the adjective Slovak probably vastly outnumber the occurrences of the word Czech. The same incorrect translations appeared also in different contexts in other sentences of the test corpus.

Morphological errors

Both languages are very similar also with regard to the number of inflected word forms derived from one lemma. This property seems to cause certain problems to both systems, as we can see in the Example 5, where both systems use an incorrect (but different) form of the same adjective. It is interesting that in this specific case the correct translation actually means no translation at all because the correct Czech and Slovak forms are identical in this context.

Example 5.

Source: *V březnu malajsijské aerolinky přišly o Boeing 777 s 239 lidmi na palubě, který se ztratil z leteckých radarů cestou z Kuala Lumpuru do Pekingu.*

Google: V marci **malajzijského** aerolinky prišli o Boeing 777 s 239 ľuďmi na palube, ktorý sa stratil z leteckých radarov cestou z Kuala Lumpur do Pekingu.

Česílko: V marci **malajsijský** aerolinky prišli o Boeing 777 s 239 človek na palube, ktorý sa stratil z leteckých radarov cestou z Kuala Lumpuru do Pekingu.

English translation: *In March, Malaysian Airlines lost Boeing 777 with 239 people on board, which got lost from air radars on the way from Kuala Lumpur to Beijing.*

Although the morphological errors have a negative influence on automatic measures like BLEU or NIST (incorrect form of a correct word influences the score to the same extent as completely incorrect word), they usually do not change the meaning of the translated sentence and the speakers of the target language can easily reconstruct their correct form and understand the translation. From this point of view both systems perform very well because the relatively low number of incorrect word forms produced by both systems doesn't reach the threshold when the sentence as a whole would be unintelligible.

Word order

Both systems follow very strictly the order of words

of source sentences. This is not surprising in the case of the RBMT system, because its simple architecture is exploiting the fact that the word order of both languages is extremely similar. As we have already mentioned in the section 3, Česílko translates word by word. The strict correspondence of the word order of source and target sentences is a bit more surprising in the case of the SMT system, whose language model is probably based on a large volume of texts with a wide variety of word order variants. Czech and Slovak languages both have very few restrictions on the order of words and thus we have supposed that the translated sentences might have an altered word order compared to the source sentences. The only difference in the order of words appeared in the sentence presented below, where the RBMT system followed the original word order strictly, while the SMT system made changes (acceptable ones) to the order of clitics.

Example 6.

Source: *Ačkoli vyšetřovatelé už jsou si jisti, že stroj se odklonil ze své trasy a zřítil se pravděpodobně po dlouhém letu nad Indickým oceánem, nevědí dodnes, co bylo příčinou nehody a nenašli ani trosky stroje.*

Google: Hoci vyšetrovatelia **sú si už** istí, že stroj sa odklonil zo svojej trasy a zrútil sa pravdepodobne po dlhom lete nad Indickým oceánom, nevedia dodnes, čo bolo príčinou nehody a nenašli ani trosky stroja.

Česílko: Bárs vyšetrovatelia **už sú si** istí, že stroj sa odklonil zo svojej trasy a zrútil sa pravdepodobne po dlhom lete nad Indickým oceánom, nevedia dodnes, čo bolo príčinou nehody a nenašli ani trosky stroja.

English translation: *Although the investigators are now sure that the plane swerved from its path and fell down probably after a long flight over the Indian ocean, they didn't find out till today what was the cause of the accident and they even didn't find any remains of the plane.*

Syntactic errors

There are no errors which could be classified as purely violating the syntax of the target language. The use of an incorrect form of direct or indirect object can be attributed to the category of morphological errors, because neither of the two systems deals with syntax directly. The RBMT system ignores syntax on the basis of the fact that both languages are syntactically very similar; the SMT system probably primarily relies on phrases discovered in large volumes of training data and thus it takes the syntactic rules into account only indirectly.

Errors in meaning

There were very few errors in incorrectly translated meaning of the source sentence into the target one. Although some SMT systems are infamous for issues related to the preservation of negated expressions, the only two examples of such errors were produced by the RBMT system in our tests. The sentence which was affected by this error to a greater extent is listed below.

No other errors in the translation of the original meaning have been encountered in our tests.

Source: *Vedení školy odpovídá na některé oficiální dotazy rodičů až po opakované urgenci a chování ředitelky Kozohorské je podle slov rodičů alibistické, nevstřícné a nezřídka arogantní.*

Google: Vedenie školy odpovedá na niektoré oficiálne otázky rodičov až po opakovanej urgencii a správanie riaditeľky Kozohorský je podľa slov rodičov alibistické, *nevstřícné a nezriedka* arogantný.

Česílko: Vedenie školy odpovedá na niektorých oficiálne otázky rodičov až po opakovanej urgencii a chovaní ředitelka Kozohorský je podľa slov rodičov alibistické, **vstřícný a zriedka** arogantní.

English translation: *The school management answers some official questions of parents after repeated reminders and the behavior of director Kozohorská is, in the words of parents, buck-passing, unresponsive and often arrogant.*

The sentence produced by Česílko has lost two negations making the behavior of the director *responsive and seldom arrogant*. This is probably caused by the fact that both the positive and negative forms have the same lemma - the negation constitutes only a small part of the morphological tag[7] and thus it may easily be forgotten or lost in the process of transfer of a Czech tag into a Slovak one (a different system of tags is used for Slovak).

6   Conclusions and further work

Although our experiment represents only the first step in systematic evaluation of machine translation results between closely related languages, it has already brought very interesting results. It has shown that contrary to a popular belief that RBMT methods are more suitable for MT of closely related languages, Google Translate outperforms the RBMT system Česílko. The similarity of source and target language apparently not only allows much simpler architecture of the RBMT system, it also improves the chances of SMT systems to generate good quality translation, although this results need further examination.

The most surprising result of our experiment is the high number of identical translations produced by both systems not only for short simple sentences, but also for some of the long ones, as well as very similar results produced for the rest of the test corpus. The minimal differences between two systems exploiting different paradigms deserve further experiments. These experiments will involve a phrase-based SMT system based on Moses (in this way we are going to guarantee that we are really comparing two different paradigms) and we will investigate its behavior on the same language pair. A second interesting experimental direction will

---

[7]Česílko exploits a positional system of morphological tags with 15 fixed positions, the negation marker occupies only one of these positions.

be the investigation whether the results for another pair of languages related not so closely as Czech and Slovak would confirm the results obtained in this experiment.

References

Lars Ahrenberg and Maria Holmqvist. 2004. Back to the Future? The Case for English-Swedish Direct Machine Translation. In *Proceedings of The Conference on Recent Advances in Scandinavian Machine Translation*. University of Uppsala.

Kemal Altintas and Ilyas Cicekli. 2002. A Machine Translation System between a Pair of Closely Related Languages. In *Proceedings of the 17th International Symposium on Computer and Information Sciences (ISCIS 2002)*, page 5. CRC Press.

Eckhard Bick and Lars Nygaard. 2007. Using Danish as a CG Interlingua: A Wide-Coverage Norwegian-English Machine Translation System. In *Proceedings of NODALIDA, Tartu*. University of Tartu.

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 *2013 Workshop on Statistical Machine Translation*, pages 1–44.

Chris Callison-Burch, Miles Osborne, and Philipp Koehn. 2006. Re-evaluating the role of BLEU in machine translation research. In *Proceedings of EACL*, pages 249–256. Association for Computational Linguistics.

Antonio M Corbi-Bellot, Mikel L Forcada, and Sergio Ortiz-Rojas. 2005. An open-source shallow-transfer machine translation engine for the Romance languages of Spain. In *Proceedings of the EAMT conference*, pages 79–86. HITEC e.V.

Helge Dyvik. 1995. Exploiting Structural Similarities in Machine Translation. *Computers and Humanities*, 28:225–245.

Jan Hajič, Jan Hric, and Vladislav Kuboň. 2000. Machine translation of very close languages. In *Proceedings of the 6th Applied Natural Language Processing Conference*, pages 7–12. Association for Computational Linguistics.

Jan Hajič, Petr Homola, and Vladislav Kuboň. 2003. A simple multilingual machine translation system. In Eduard Hovy and Elliott Macklovitch, editors, *Proceedings of the MT Summit IX*, pages 157–164, New Orleans, USA. AMTA.

Petr Homola and Vladislav Kuboň. 2008. A method of hybrid MT for related languages. In *Proceedings of the IIS*, pages 269–278. Academic Publishing House EXIT.

Petr Homola and Jernej Vičič. 2010. Combining MT Systems Effectively. In *Proceedings of the 23th International Florida-Artificial-Intelligence-Research-Society Conference (FLAIRS 2010)*, pages 198–203, Daytona Beach, Florida, USA. Florida {AI} Research Society, Florida AI Research Society.

David Kolovratnık, Natalia Klyueva, and Ondrej Bojar. 2009. Statistical machine translation between related and unrelated languages. In *Proceedings of the Conference on Theory and Practice on Information Technologies*, pages 31–36.

NIST. 2009. NIST 2009 Open Machine Translation Evaluation (MT09). Technical report, NIST.

Karel Oliva. 1989. *A Parser for Czech Implemented in Systems Q*. MFF UK Prague.

Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. 2001. BLEU: a method for automatic evaluation of machine translation. Technical report, IBM.

Kevin P Scannell. 2006. Machine translation for closely related language pairs. In *Proceedings of the Workshop Strategies for developing machine translation for minority languages*, pages 103–109. Genoa, Italy.

Francis M Tyers, Linda Wiechetek, and Trond Trosterud. 2009. Developing prototypes for machine translation between two Sámi languages. In *Proceedings of EAMT*. HITEC e.V.

Jernej Vičič. 2008. Rapid development of data for shallow transfer RBMT translation systems for highly inflective languages. In *Language technologies: proceedings of the conference*, pages 98–103. Institut Jožef Stefan, Ljubljana.

Per Weijnitz, Eva Forsbom, Ebba Gustavii, Eva Pettersson, and Jorg Tiedemann. 2004. MT goes farming: Comparing two machine translation approaches on a new domain. In *LREC*, pages 1–4.