

# Speeding up the Implementation Process of a Shallow Transfer Machine Translation System

**Jernej Vičič**

University of Primorska, PINT  
Glagoljaška 8, Koper Slovenia  
jernej.vicic@upr.si

**Petr Homola**

Institute of Formal and Applied Linguistics  
Malostranské náměstí 25, Praha, Czech Republic  
homola@ufal.mff.cuni.cz

## Abstract

The article presents an attempt to automate all data creation processes of a rule-based shallow-transfer machine translation system. The presented methods were tested on two fully functional translation systems Slovenian-Serbian and Slovenian-Macedonian. An extensive range of evaluation tests was performed to assess the applicability of the methods.

## 1 Introduction and problem statement

Several methods that automate some parts of the shallow transfer Rule Based Machine Translation (RBMT) system construction have been presented and are even used as part of the construction toolkits like Apertium (Corbi-Bellot et al., 2005), which is a widely used architecture for creating machine translation systems between related languages. (Hajič et al., 2000), (Hajič et al., 2003) and (Corbi-Bellot et al., 2005) suggest using an architecture similar to the one presented in Figure 1. All methods and materials discussed in this paper were tested on a fully functional machine translation system based on Apertium.

The construction of a machine translation system for a new language pair falls into roughly two categories:

- A job of manual dictionary and rule construction for Rule-Based Machine Translation (RBMT) system construction.
- Automatic machine translation system construction in case of corpus-based machine construction systems such as Statistical Machine Translation (SMT) (Brown et al., 1993;

Och and Ney, 2003) or Example-Based Machine Translation (EBMT) (Nagao, 1984; Hutchins, 2005).

The SMT seems like a perfect choice as some of the best performing machine translation systems are based on the SMT technologies (NIST, 2006), but it has a few drawbacks that cannot be ignored; the SMT systems, to be efficient, require huge amount of parallel text (Och, 2006) that is available only for a few of the widely used languages like English, Spanish, French, Arabic, etc. The morphologically rich and highly inflective languages like the pairs presented in this paper (Slovenian paired with Serbian and Macedonian) present an even bigger problem.

One of the most appealing reasons for using an RBMT machine translation system is the ability for the experts of the field to further refine the results of the automatically produced data.

An attempt to fully automatise the construction of all the data for a fully functional shallow transfer RBMT system has been presented in (Vičič, 2008). Parts of the creation process have been addressed by several authors such as automated monolingual dictionary extraction (Forsberg et al., 2006); support for agglutinative languages (Beesley and Karttunen, 2000); Part Of Speech (POS) tagger training (Sánchez-Martínez et al., 2008; Halácsy et al., 2007; Brants, 2000); automatic induction of shallow-transfer rules (Sanchez-Martínez and Forcada, 2009); automatic extraction of bilingual dictionaries (Caseli et al., 2008). Some of these technologies are used in this paper along with newly developed methods.

The rest of the article is organised as follows: The architecture of a typical Shallow Transfer RBMT system is presented in Section 2. The Section 3

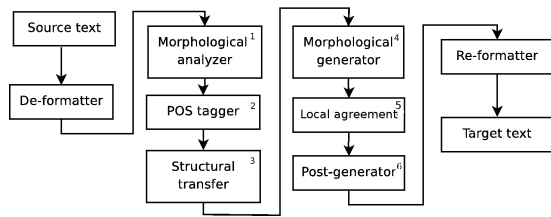


Figure 1: The modules of a typical shallow transfer translation system. The systems (Corbi-Bellot et al., 2005; Hajič et al., 2003) follow this design. An addition of the original architecture is the local agreement module tagged as number 6.

presents the addressed language pairs, follows a presentation of the used methods in Section 4. The evaluation methodology with results is presented in Section 5, the article concludes with the discussion in Section 6.

## 2 Architecture of a typical shallow transfer RBMT system

The machine translation system used in the experiments described in this paper is based on Apertium, an open-source RBMT toolkit. Apertium is an open-source machine translation platform, initially aimed at related-language pairs but recently expanded to deal with more divergent language pairs (such as English-Catalan). The shallow-transfer paradigm of the toolkit is best suited for related languages as the architecture does not provide the means for deep parsing which can lead to problems especially in the more divergent language pairs. All these properties make Apertium a perfect choice in a cost effective development of a machine translation system for similar languages. The basic architecture of Apertium system is presented in Figure 1.

The monolingual dictionaries are used in the morphological parsing of the source text by the morphological analyser module, the rectangle 1 in Figure 1, and in the generation of the translation text in the target language by the morphological generator module, the rectangle 4 in Figure 1. The bilingual dictionary is used for word-by-word translation, in our case the translation is based on lemmata. The shallow transfer rules are used to address local syntactic and morphological rules such as local word agreement and local word reordering. The module using the bilingual dictionary and the shallow transfer rules is the structural transfer

module, the rectangle 3 in Figure 1.

The finite state local agreement rules of the target language are used as a refinement method in the post-processing phase to eliminate errors produced by shallow transfer rules in the transfer phase by the local agreement module, the rectangle 5 in Figure 1.

Each module from the list was addressed by applying either a known method or by introducing a new method. The methods are presented in more detail in Section 4. A fully functional system was constructed using the presented methods and overall performance of the whole system was evaluated.

## 3 Chosen languages and description of the translation systems

The language pairs used in this experiment were Slovenian-Serbian (SL-SR) and Slovenian-Macedonian (SL-MK). All three languages belong to the group of Southern-Slavic languages that were mostly spoken on the territory of the former Yugoslavia. Now these languages are mostly spoken on the territories of Slovenia, Serbia and Macedonia respectively. The economies of the nations where these languages are spoken are closely connected and younger generations, the post-Yugoslavia breakage generations, have difficulties in mutual communication, so there is a big interest in construction of such translation system. Slavic languages have rich conjugation and most of them (except for Bulgarian and Macedonian) rich nominal declension. Slovenian and Serbian languages do not use articles, the category of the definiteness is expressed by the word order which is extremely flexible although similar among the group. The Macedonian language uses articles but it does not have the fixed word order.

Although the languages are related, the described properties of the language pair demand morpho-syntactical analysis of the source text and later morpho-syntactical synthesis of the target text.

Figure 2 shows a sentence in all three languages. The noun *okno* changes gender from neutral in Slovenian to masculine in Serbian and Macedonian. The case of the noun and the adjacent adjective *veliko* - *big* changes from nominative in Slovenian and Serbian to *non-applicable* in Macedonian.

Veliko okno je odprto.  
Veliki prozor je otvoren.  
Голям прозор е отворен.

Figure 2: An example sentence in all three languages (Slovenian, Serbian, Macedonian). *A big window is open.*

### 3.1 Available resources

The most important resource used in this experiment was the multilingual parallel and comparative corpus MULTEXT-EAST (Dimitrova et al., 1998). Part of the corpus (randomly selected sentences) was used for training, the rest was used for testing purposes. The same sentences were selected for each language.

### 3.2 Description of the systems

The system using Slovenian-Serbian language pair was constructed during the method development process. The methods presented in this paper were checked through several iterations (the systematic errors were corrected and included in the basic framework). This language pair was used to check the quality of the presented methods on a fully functional translation system. Both languages are highly inflective, morphologically and derivationally rich. Although these languages are related, the high degree of inflection in both languages of the pair still demands the morphological analysis of the source language and later morphological synthesis of the target language.

The Slovenian-Macedonian system was constructed to evaluate the applicability of the methods presented in Section 4 on a new language pair of related languages and to test how quickly a new system can be constructed. This test was made in order to check if the methods are language independent.

The similarity of this language pair is smaller in comparison to the first language pair, the differences are shown in Section 3. The system was constructed from scratch in just two days by a single person on an affordable computer<sup>1</sup>.

<sup>1</sup>A notebook computer with 2 GB of RAM and an Intel Core 2 duo processor.

## 4 Methodology

The modules presented in Figure 1 and numbered with numbers 1 through 5 require linguistic data (monolingual dictionaries, bilingual dictionaries, translation rules, etc.). The next subsections describe methods that automate the linguistic data creation process for each module.

### 4.1 Monolingual source and target dictionary creation

If we take an example from English; the transformation of the word *walk* into *walked* can be achieved by a morphological transformation rule (for the past tense). A variation of the same rule would be used for the irregular word *sleep*, changing into *slept*. For the languages that employ inflectional morphology where words are composed of a number of morphemes concatenated together; the morphemes include the stem plus prefixes and suffixes such as the majority of European languages, different forms of the same word are produced by changing the prefix and suffix of the word. Thus, *slept* can be derived from *sleep* by changing the suffix *-ep* to the suffix *-pt*. The same phenomenon, but to a much greater extent, occurs in highly inflectional languages.

#### 4.1.1 Paradigm creation

The words were grouped into paradigms in order to deal with multiple word forms as both Slovenian and Serbian are highly inflectional languages. Each paradigm is represented by:

- a typical lemma; the lemma the paradigm was constructed from,
- a stem; the longest common prefix of all words in the lemma,
- a set of all words split into stems, suffixes and Morpho-Syntactic Descriptors (MSDs) (Erjavec, 2004).

An example is shown in Figure 3.

The annotated lexicons, lists of unique words with lemma descriptor and MSD, were extracted from corpus for both languages and paradigms were constructed using the presented method.

All of the word forms of a lemma present in the corpus are grouped into a class represented by lemma. A paradigm is constructed from each class KKKK for each lemma. Two paradigms are joined together if the lemmata of both paradigms

```

lemma: cerkev
stem: cerk
example entries:
word form: cerkev
suffix: ev
MSD: noun+feminine+singular+nominative
word form: cerkvah
suffix: vah
MSD: noun+feminine+plural+locative

```

Figure 3: A part of a paradigm *cerkev* - church. Lemma: cerkev, stem: cerk, two word forms |emphcerkev and *cerkvah*

have the same POS tag and if the entries, pairs of suffix and MSD, of one paradigm present a complete subset of the compared paradigm. The complexity of this process increases linearly as the number of lemmata in paradigms increases by joining paradigms. The information about all lemmata that generated the paradigm is stored in a list enabling easy lookup. The monolingual source and target dictionaries were constructed using joined paradigms resulting in a lexicon that was roughly 20 times larger the original lexicon.

## 4.2 Bilingual translation dictionary creation

An SMT word-to-word model (Brown et al., 1993; Och and Ney, 2003) was trained on the parallel, sentence aligned corpus. The corpus was lemmatized and POS tagged.

The experiment involved richly inflected languages where lemmatization of a text greatly decreases the number of unique tokens. Table 1 shows the difference in number of word forms for the same corpus (Dimitrova et al., 1998) in five languages; three rich inflectional Slavic languages: Slovenian, Serbian, Czech along with English and Estonian for reference.

Table 1: Number of lemmata in corpus MULTEXT-EAST (Dimitrova et al., 1998)

language	nr. of unq. words	lemmata	ratio <sup>2</sup>
Slovenian	20,923	7,895	2.65
Serbian	21,505	8,392	2.56
Czech	22,273	9,060	2.46
English	11,078	7,020	1.58
Estonian	18,853	8,679	2.17

The reduction of search space increases the accuracy of the model (the word-by-word trans-

lation model). This result is not surprising, but a lot of information about the word form was lost in the process. The lemmata alignment ensures much better alignment performance due to the search space reduction as described in Table 1. The words from the monolingual dictionaries are aligned to the translations (bilingual lemmata pairs) through paradigms that retain the information about the included lemmata, see Section 4.1.1.

The bilingual parallel annotated corpus (Dimitrova et al., 1998) comprises original text with additional information in the form of XML tags according to the TEI-P4 (Consortium, 2007) and the EAGLES (Leech and Wilson, 1996) guidelines.

Each word is represented by the *lemma* (lemma of the word), *ana* (morphosyntactical description - MSD (Erjavec, 2004)) and the word form used in the corpus. Only the lemma and POS of each word are extracted from the corpus for this task, leaving parallel sentences in lemmatised form with the POS tag. Figure 4 shows the prepared data. The model was trained using GIZA++ tool (Och and Ney, 2003) on the prepared data aligning the words in one direction using a minimum number of tokens in the corpus and a probability threshold.

```

priti_V biti_V do_S podrt_A drevo_N ,
o_S kateri_P on_P biti_V praviti_V .

```

Figure 4: Prepared data: a selected part of the lemmata and POS pairs from the corpus

### 4.2.1 Extension of the bilingual dictionary

This method extends the number of entries in bilingual dictionary and monolingual dictionaries. The methods described in the Sections 4.1 and 4.2 do not guarantee full coverage of the monolingual dictionaries by the bilingual dictionary. The bilingual dictionary constructed using the method presented in Section 4.2 comprised of around 3800 lemmata. The method presented in this section was used to construct a bigger bilingual dictionary. This method exploits the lexical similarity of the observed languages.

The description of the basic algorithm of the method: A new lemma pair is created for each entry of the source and target monolingual dictionary that has no translation entry already present in the bilingual dictionary. The translation entry in the bilingual dictionary is composed of the

same lemma in source and target part. A suitable POS descriptor, the POS tag of the MSD (Erjavec, 2004) descriptor, is added to the available lemma and also copied to the counter part of the bilingual pair. The first example in Figure 5 shows an entry in the source, Slovenian, monolingual dictionary. The lemma *list* had no translation in the bilingual dictionary. The second example in Figure 5 was added to the bilingual dictionary Slovenian-Serbian, the entry translates Slovenian lemma *list* to the target Serbian lemma *list*.

A new entry is added to the target monolingual dictionary if the lemma with no translation was found in the source monolingual dictionary and vice-versa. A suitable paradigm is searched in the target monolingual dictionary by selecting a paradigm with the longest applicable suffix and suitable (same) POS descriptor. The third example in Figure 5 shows the entry added to the target dictionary; the lemma *list* was added to the dictionary and joined to the paradigm "*um/\_\_\_n*". A transliteration was needed for the Slovenian to Macedonian bilingual dictionary as Macedonian uses only Cyrillic script, so source entries were translated from Latin to Cyrillic and the target entries were translated from Cyrillic to Latin. The example in Figure 6 shows such an entry.

#### 4.3 Shallow transfer finite-state rules induction

The shallow transfer, finite-state type rules, were constructed using available software from the Apertium toolkit. The software is based on the technologies presented in (Sanchez-Martinez and Forcada, 2009). The basic idea of the process is using statistical methods to construct templates from bilingual aligned corpus. These templates are later translated into finite-state rules in the Apertium format.

#### 4.4 Automatic induction of local agreement rules

The automatic induction of the local agreement rules produces the same format of the rules as the method described in (Sanchez-Martinez and Forcada, 2009), but the method is limited to the discovery of local agreement. The requirements for the method are much simpler, just a morphologically annotated corpus of the source language. Trigrams and bigrams with morphological descriptions were extracted from the source language part of the corpus. The corpus used as training data was

(Dimitrova et al., 1998), which was hand checked for errors in morphosyntactic tags. The source language used in our experiment was Slovenian language although the same method could be used for other languages as the method is not language specific. Each bigram and trigram was checked for agreement among tags of different words, the tags and their positions were arbitrary. If any agreements were found, a candidate for a rule was stored. The POS tags of the source bigram or trigram represent the pattern part of the rule. The action part of the rule is constructed from all the morphosyntactic tags with agreement information. The rule candidates were grouped according to the pattern and action definitions, each group with a predefined number of candidates was chosen as a valid rule.

## 5 Evaluation methodology and results

The methods presented in Section 4 are targeted at the construction of machine translation systems for related, morphologically rich languages. The experiments are aimed at showing the quality of automatically generated data on a fully functional translation system and also on the usability of presented methods for rapid development of a translation system for a new language pair.

Two fully functional translation systems were constructed and evaluated in this experiment:

1. SL-SR, Slovenian to Serbian translation system
2. SL-MK, Slovenian to Macedonian translation system

### 5.1 Description of the evaluation metrics

The evaluation of the translations was performed in three parts, each part is further described below:

1. The automatic objective evaluation using the METEOR (Lavie and Agarwal, 2007) metric.
2. The non-automatic evaluation by counting the number of edits needed to produce a correct target sentence from an automatically translated sentence.
3. The non-automatic subjective evaluation following (LDC, 2005) guidelines.

Preliminary evaluation plans included the evaluation using the BLEU (Papineni et al., 2001) metric as it is one of the most used machine translation evaluation metrics. Many authors agree that

Slovenian monolingual dictionary:

```
<e lm="list"><i>list</i><par n="žvenket/___n"/></e>
```

Slovenian - Serbian bilingual dictionary:

```
<e><p><l>list<s n="n"/><s n="m"/></l>
<r>list<s n="n"/><s n="m"/></r></p></e>
```

Serbian monolingual dictionary:

```
<e lm="list"><i>list</i><par n="um/___n"/></e>
```

Figure 5: The extension of the bilingual corpus. First example shows a Slovenian lemma *list*, *leaf*, and the additional entries in the Serbian monolingual and SL-SR bilingual dictionaries

Slovenian monolingual dictionary:

```
<e lm="list"><i>list</i><par n="žvenket/..n"/></e>
```

SL-MK bilingual dictionary:

```
<e><p><l>list<s n="n"/></l><r>лист<s n="n"/></r></p></e>
```

Macedonian monolingual dictionary:

```
<e lm="лист"><i>лист</i><par n="капиталист/..n"/></e>
```

Figure 6: The extension of the bilingual corpus. First example shows a Slovenian lemma *list* *leaf* and the additional entries in the Macedonian monolingual and SL-MK bilingual dictionaries. The Macedonian entries are in Cyrillic.

BLEU metric systematically penalises RBMT systems (Callison-Burch et al., 2006; Labaka et al., 2007) and it is not suited for highly inflective languages. The evaluation using the BLEU metric was not performed. Authors of METEOR state that their system fixes most of the problems encountered using the BLEU metric; they state that METEOR correlates highly with human judgement. Unfortunately METEOR did not support our language pair, additional software had to be written. The bilingual parallel corpus (Dimitrova et al., 1998) was used in automatic evaluation of translations. Five-fold cross-validation was used as the method for estimating the generalisation error as it is most suitable for small data sets. The corpus was divided into five parts, each part consisting of roughly 1,700 sentences. The evaluation consisted of selecting one part of the corpus as the test set and remaining four parts as the training set. The translation system was constructed according to the methodology presented in Section 4 using the selected training set. The evaluated values in each fold and the average final values are presented.

### 5.1.1 Automatic objective evaluation using the METEOR metric

The publicly available implementation of the METEOR metric (Lavie and Agarwal, 2007) version 0.6 was used. The METEOR metric uses stemming mechanism as one of the algorithms that enhance correlation between the METEOR metric and human evaluation for highly inflectional lan-

guages. The stemming mechanism that is a side-product of the described translation system was used. Results are presented in Figure 7.

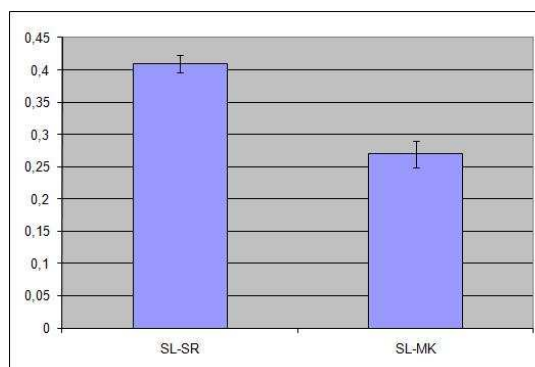


Figure 7: The METEOR metric scores. The evaluation was done using 5-fold cross validation. The values in the Figure represent the average values of 5 folds with standard deviation.

### 5.1.2 Non-automatic evaluation using edit distance

The weighted Levenshtein edit-distance (Levenshtein, 1965) or more commonly known as Word Error Rate (WER) was used to count the number of edits needed to produce a correct target sentence from automatically translated sentence. This procedure shows how much work has to be done to produce a good translation. The metric roughly reflects the complexity of the post-editing task. The evaluation comprised of selecting 200 sen-

tences from the test data, translating these sentences using the translation system and manually counting the number of words that had to be changed in order to obtain a perfect translation. By perfect translation we mean a translation that is syntactically correct and expresses the same meaning as the source sentence.

The evaluation and the results presented in the Figure 8 present the WRR, the Word Recognition rate ( $1 - \text{WER}$ ), which presents the performance of the translation system instead of the error.

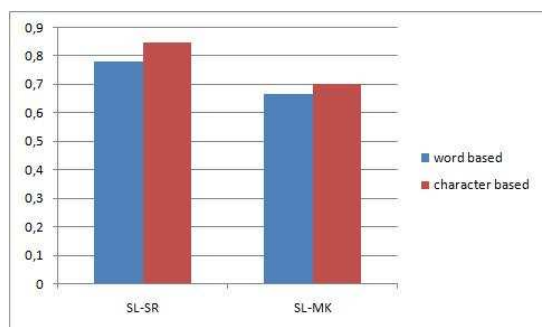


Figure 8: The evaluation results using the Word Recognition Rate metric.

### 5.1.3 Non-automatic subjective evaluation following (LDC, 2005) guidelines

Subjective manual evaluation of translation quality was performed according to the annual NIST Machine Translation Evaluation Workshop by the Linguistic Data Consortium guidelines. The most widely used methodology when manually evaluating MT is to assign values from two five-point scales representation fluency and adequacy. These scales were developed for the annual NIST Machine Translation Evaluation Workshop by the Linguistic Data Consortium (LDC, 2005).

The five point scale for adequacy indicates how much of the meaning expressed in the reference translation is also expressed in a hypothesis translation: 5 = All, 4 = Most, 3 = Much, 2 = Little, 1 = None. The second five-point scale indicates how fluent the translation is. It expresses whether the translation is syntactically well formed. When translating into Serbian the values correspond to: 5 = Flawless translation, 4 = Good target language, 3 = Non-native target language, 2 = Disfluent target language, 1 = Incomprehensible text. Separate scales for fluency and adequacy were developed under the assumption that a translation might

be disfluent but contain all the information from the source. Four independent evaluators (two native speakers) evaluated sets of 100 sentences using this methodology. The results are presented in Figure 9. The standard deviation shows the degree of inter-rater agreement. As shown on the Figure 9, the inter-rater agreement was satisfactory (low standard deviation).

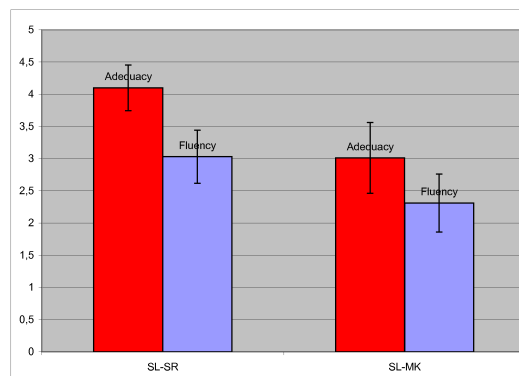


Figure 9: Evaluation results using (LDC, 2005) guidelines. Average values of four independent evaluations show high scores for adequacy and lower values for fluency.

## 6 Discussion and further work

The agreement among all three evaluation methods is quite high, which shows that the results of the evaluation process are valid. The translation quality of the Slovenian-Serbian translation system is higher than the quality of the translation system Slovenian-Macedonian. This can be attributed to the fact that the similarity of the later language pair is lower than the first one. The basic differences are the lack of nominal declension in Macedonian in comparison to the other two languages which share the same 6 cases. The other difference is the expression of definiteness using the articles for the Macedonian language and using word order for the other two languages.

The automatically generated linguistic data is far from perfect and additional manual labour will have to be executed in order to obtain better translation quality. The experiment showed that the methods can be adopted for a number of relatively close language pairs. Our intention is to construct translation systems among all the MULTTEXT-EAST (Dimitrova et al., 1998) languages including the recent additions such as Macedonian and Croatian. One of the experiments will be the con-

struction of a pan-Yugoslav translation system that would cover all the ex-Yugoslavian languages.

## References

- Beesley, K. and L. Karttunen. 2000. Finite-state non-concatenative morphotactics. In *Proceedings of the Workshop of the ACL-SIGCP*, pages 1–12.
- Brants, Thorsten. 2000. TnT—a statistical part-of-speech tagger. In *Proceedings of the 6th Applied NLP Conference*. Seattle, WA.
- Brown, Peter F., Stephen A. Della Pietra, Vincent J. Della Pietra, and Robert L. Mercer. 1993. The mathematics of statistical machine translation: parameter estimation. *Computational linguistics*, 19(2):163–311.
- Callison-Burch, Chris, Miles Osborne, and Philipp Koehn. 2006. Re-evaluating the role of Bleu in machine translation research. In *Proceedings of EACL*.
- Caseli, Helena M., Maria das Graças V. Nunes, and Mikel L. Forcada. 2008. From free shallow monolingual resources to machine translation systems easing the task. In *Proceedings of the Workshop of the ACL-SIGCP*, pages 41–48.
- Consortium, TEI. 2007. TEI P5: Guidelines for Electronic Text Encoding and Interchange. Technical report, TEI consortium.
- Corbi-Bellot, Antonio M., Mikel L. Forcada, Sergio Ortiz-Rojas, Juan Antonio Prez-Ortiz, Gemma Ramirez-Sanchez, Felipe Sanchez-Martinez, Inaki Alegria, Aingeru Mayor, and Kepa Sarasola. 2005. An open-source shallow-transfer machine translation engine for the Romance languages of Spain. In *Proceedings EAMT conference*, pages 79–86, May.
- Dimitrova, Ludmila, Nancy Ide, Vladimir Petkevič, Tomaž Erjavec, Heiki Jaan Kaalep, and Dan Tufis. 1998. Multext-East: Parallel and Comparable Corpora and Lexicons for Six Central and Eastern European Languages. In *COLING-ACL*, pages 315–319.
- Erjavec, Tomaž. 2004. MULTTEXT-East Version 3: Multilingual Morphosyntactic Specifications, Lexicons and Corpora. In *Proc. of the Fourth Intl. Conf. on Language Resources and Evaluation, LREC'04*.
- Forsberg, M., H. Hammarström, and A. Ranta. 2006. Morphological lexicon extraction from raw text data. In *Proceedings FinTAL 2006, LNAI 4139*, pages 488–499.
- Hajič, J., J. Hric, and V. Kuboň. 2000. Machine translation of very close languages. In *Proceedings of the 6th Applied Natural Language Processing Conference*.
- Hajič, Jan, Petr Homola, and Vladislav Kuboň. 2003. A simple multilingual machine translation system. In *Proceedings of the MT Summit IX*, New Orleans.
- Halácsy, Kornai, and Oravecz. 2007. HunPos—an open source trigram tagger. In *Proceedings of the ACL 2007 Demo and Poster Sessions*, pages 209–212.
- Hutchins, John. 2005. Towards a definition of example-based machine translation. In *MT Summit X, Proceedings of Workshop on EBMT*, pages 63–70.
- Labaka, Gorka, Nicholas Stroppa, Andy Way, and Kepa Sarasola. 2007. Comparing rule-based and data-driven approaches to Spanish-to-Basque machine translation. In *Proceedings of the Machine Translation Summit XI*, pages 41–48.
- Lavie, A. and A. Agarwal. 2007. METEOR: An Automatic Metric for MT Evaluation with High Levels of Correlation with Human Judgments. In *Proceedings of Workshop on SMT at the ACL conference*.
- LDC. 2005. Linguistic data annotation specification: Assessment of fluency and adequacy in translations. Technical report, LDC.
- Leech, GN and A. Wilson. 1996. Eagles recommendations for the morphosyntactic annotation of corpora. Technical report, ILC-CNR, Pisa.
- Levenshtein, V. 1965. Binary codes capable of correcting deletions, insertions and reversals. *Doklady Akademii Nauk*, pages 845–848.
- Nagao, Makoto. 1984. A framework of a mechanical translation between Japanese and English by analogy principle. *Artificial and Human Intelligence*.
- NIST. 2006. Nist 2006 machine translation evaluation official results.
- Och, Franz Josef and Hermann Ney. 2003. A systematic comparison of various statistical alignment models. *Computational linguistics*, 29:19–51.
- Och, Franz Josef. 2006. Challenges in Machine Translation. In *Proceedings of ISCSLP*.
- Papineni, Kishore, Salim Roukos, Todd Ward, and Wei-Jing Zhu. 2001. Bleu: a method for automatic evaluation of machine translation. Technical report, IBM.
- Sanchez-Martinez, Felipe and Mikel L. Forcada. 2009. Inferring shallow-transfer machine translation rules from small parallel corpora. *Journal of Artificial Intelligence Research*, 34:605–635.
- Sánchez-Martínez, Felipe, Juan Antonio Pérez-Ortiz, and Mikel L. Forcada. 2008. Using target-language information to train part-of-speech taggers for machine translation. *Machine Translation*, 22(1-2):29–66. DOI: 10.1007/s10590-008-9044-3.
- Vičič, Jernej. 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.