# Computing transfer score in Example-Based Machine Translation

Rafał Jaworski

Adam Mickiewicz University
Poznań, Poland
rjawor@amu.edu.pl

**Abstract.** This paper presents an idea in Example-Based Machine Translation - computing the transfer score for each produced translation. When an EBMT system finds an example in the translation memory, it tries to modify the sentence in order to produce the best possible translation of the input sentence. The user of the system, however, is unable to judge the quality of the translation. This problem can be solved by providing the user with a percentage score for each translated sentence.
The idea to base transfer score computation on the similarity between the input sentence and the example is not sufficient. Real-life examples show that the transfer process is as likely to go well with a bad translation memory example as to fail with a good example.
This paper describes a method of computing transfer score strictly associated with the transfer process. The transfer score is inversely proportional to the number of linguistic operations executed on the example target sentence. The paper ends with an evaluation of the suggested method.

## 1 Introduction

During the latest studies on EBMT (e.g. [1], [2], [3]), a serious problem with the development of usable EBMT systems emerged. The usability of a newly created EBMT system cannot be assessed based on automatic machine translation metrics, such as BLEU ([4]). There are two reasons behind it. Firstly, the automatic translation metrics were created rather for RBMT systems and the scores they produce do not reflect the nature of EBMT translations. Secondly, it is hard to say whether a system scoring well in BLEU (or any automatic metric) can be used in real-life situations.
As the issue of usability has been raised, a definition of usability of an EBMT system must be provided. This paper describes the implementation of an EBMT system as a translation-aide tool. CAT (Computer-Aided Translation) systems are not intended to replace a human translator at all stages of the translation process. Their role is limited merely to suggesting a possible translation of a sentence to a professional translator who performs post-editing. In that sense, a usable EBMT system is a system producing such translations that are easy to correct and based on which a good translation can be obtained with minimum

effort. Therefore, the translation score computation process described in this paper serves as a measure of usability of translations produced by an EBMT system. The translation score metric is an answer to the drawbacks of automatic machine translation metrics like BLEU.
Section 2 of this paper describes the architecture of an EBMT system in which the transfer score computation mechanism was embedded. Section 3 outlines the transfer score computation mechanism itself. Section 4 compares the transfer score metric to other automatic machine translation metrics. It also discusses the experiment of comparing the transfer score metric with human judgment of usability of translations. The final section presents the conclusions.

## 2 EBMT system architecture

The system basic architecture resembles the architectures of other EBMT implementations. Like the EBMT system designed at the Chinese Academy of Science ([5]), it consists of two basic modules: Example Matcher and Transferer. The former is responsible for finding an example best suited for the input sentence in the translation memory. The latter tries to modify the example's target sentence so that it can be returned as translation of the input sentence.

### 2.1 Word substitution

The Transferer module performs operations to produce the translation of the input sentence. The crucial one is word substitution. The mechanism of this substitution can be explained using the following example:

INPUT SENTENCE (in Polish): "Uwzględniając Traktat ustanawiający Parlament Europejski". (in English: Having regard to the Treaty establishing the European Parliament).

Example from the translation memory:
SOURCE SENTENCE (in Polish): "Uwzględniając Traktat ustanawiający Wspólnotę Europejską".
TARGET SENTENCE (in English): "Having regard to the Treaty establishing the European Community."

The first operation is to check the resemblance of the input and source sentences. The result of this operation is presented in Figure 1. Each word of the input sentence is assigned to a word in the source sentence. The assignments are established using a monolingual dictionary. Solid lines represent equal words, while dotted lines represent different word forms within the same lexeme. The only word left unassigned in the input sentence is *"Parlament"* (Parliament), which does not have a match in the source sentence. Similarly, the word *"Wspólnotę"* (Community) in the source sentence does not have a match in the input


**Fig. 1.** The resemblance of the input and source sentences

sentence. At this point a significant decision is made. The word *"Wspólnotę"* in the example is substituted with the word *"Parlament"* from the input sentence. The consequence of this decision is that the example target sentence, which will be used to produce the translation, must also be changed (in response to the change in the source sentence). In order to specify what modifications the target sentence must undergo, the resemblance of the source and target sentences must be checked. This operation corresponds to the operations in [6]. Here, however, it is performed during the transfer, not during the preparation of the translation memory. The operation is done using a bilingual dictionary. The results are presented in Figure 2. The dotted lines represent the alleged correspondence of

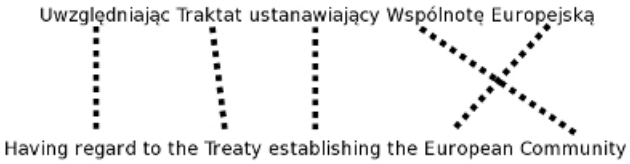


**Fig. 2.** The resemblance of the source and target sentences

Polish and English words (words which can be each other's translations are connected with these lines). The most significant information in the above diagram is the correspondence of the word *"Wspólnotę"* with the word "Community". It allows us to find the spot for the word *"Parlament"* from the input sentence. The Transferer module will put the word *"Parlament"* from the input sentence into the example's target sentence in the place of the word "Community". The result is as follows:

"Having regard to the Treaty establishing the European *Parlament*".

The final step is translating the word *"Parlament"* into English using a dictionary. This operation produces the final translation output:

"Having regard to the Treaty establishing the European Parliament".

### 2.2 NE recognition and substitution

A Named Entity in this EBMT system is defined as one or more words of a sentence having special meaning and to which special translation rules must be applied. Named Entity recognition plays an important role in the process of Machine Translation ([7]). Named Entities often carry the most important information in a sentence. At the same time, Named Entities are prone to translation errors because of their irregularity. Hence, dealing with Named Entities during the process of translation can considerably improve translation quality ([7]).
Here, Named Entities are used in a substitution mechanism similar to word substitution. There are, however, two main differences between the two mechanisms. First, words are extracted from the sentences by tokenization, while recognizing Named Entities is done by a special Named Entity recognition mechanism. The other difference lies in the process of translation - words are translated using dictionaries while Named Entity translation requires special Named Entity translation rules.
Named Entity recognition and translation is handled using rules written in a formalism called NERT (full specification of the formalism is available in [7]).
The following example illustrates the substitution of Named Entities during the process of transfer.

INPUT SENTENCE (in Polish): "Przepis Dz.U. z 12.03.2001 NR 4, poz. 17 brzmi następująco" (in English: "The regulation in Journal of Laws of 2001/03/12 No. 4, item 17 states the following").

Example from the translation memory:
SOURCE SENTENCE (in Polish): "Przepis Dz.U. z 17.01.1997 NR 8, poz. 2 brzmi następująco" (changed reference)
TARGET SENTENCE (in English): "The regulation in Journal of Laws of 1997/01/17 No. 8, item 2 states the following"

As with word substitution, the first step is to check the resemblance of the input and source sentences. During the process, Named Entities are recognised. The results of resemblance check are presented in Figure 3. The only difference between the two sentences is the target of the reference to the Journal of Laws. Therefore, the reference in the example will be substituted with the reference from input sentence. The information on resemblance of the source and input sentences is needed to complete the substitution. It is presented in Figure 4. The reference 'Dz.U. z 12.03.2001 NR 4, poz. 17' from the input sentence is to be put into the target sentence in place of the reference 'Journal of Laws of 1997/01/17 No. 8, item 2'. The reference is also translated into English. The resulting translation is:

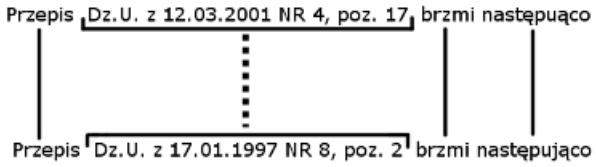


**Fig. 3.** The resemblance of the input and source sentences while substituting Named Entities

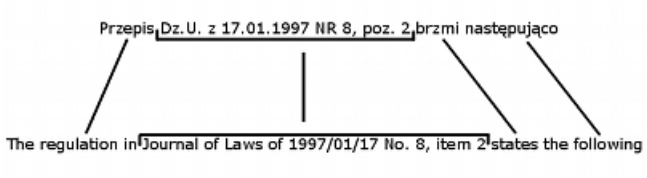


**Fig. 4.** The resemblance of the source and target sentences while substituting Named Entities

"The regulation in Journal of Laws of 2001/03/12 No. 4, item 17 states the following".

## 3 Transfer score computation mechanism

### 3.1 General idea

The idea of computing the transfer score in an EBMT system may be found in the ANTRA system [8]. However, the transfer score computation mechanism used in this system has proven not to reliable enough. Hence a different method of computing translation score has been developed.

### 3.2 Transfer score computation algorithm

The transfer score computation is done by the Transferer module during the process of transfer. Hence, it is strictly connected with the transfer. The process consists of 3 steps:

1. Finding discrepancies between input and source sentences.
2. Imposing "penalties" for discrepancies.
3. Compensating "penalties" previously imposed for discrepancies (executed while transferring words and Named Entities from the input to the target sentence).

The "penalties" play a major role in score computation. They are real, positive numbers, representing discrepancies between the source and input sentences. These discrepancies are identified in step 1 of the score computation process. The more discrepancies, the less chance for the transfer process to produce a good translation, hence more penalties. The following table presents penalties for given discrepancies (the following abbreviations has been used: IS - input sentence, SS - source sentence):

**Table 1.** Penalties imposed for discrepancies between input and source sentences.

| **Discrepancy** | **Penalty value** |
|---|---|
| Dictionary correspondence (not identity) of words | 0.5 |
| Type correspondence (not identity) of NE | 0.5 |
| Word from IS missing in SS | 1.0 |
| Word from SS missing in IS | 1.0 |
| NE from IS missing in SS | 1.5 |
| NE from SS missing in IS | 1.5 |
| Inversion of pair of words | 0.5 |
| Missing punctuation mark | 0.25 |

The second step of computing the transfer score is taken during transferring words and Named Entities from the input to the target sentence. For each successful transfer, the penalty imposed previously is partially compensated for. The table below shows the compensations for each successful transfer operation: Note that when a transfer of words is possible, a 2.0 penalty must have been imposed during checking of resemblance between IS and SS (The tranferred word from IS must have been missing in SS. Also, the word from SS, in whose place the tranferred word was put, must have been missing in IS). The compensation leaves a penalty of 0.5 (a measure of uncertainty, as to whether the transfer of the word produced a correct translation). As for NE substitution, the compensation leaves only 0.1 penalty, as transfer of NE's is more reliable (once a NE is recognized, its transfer is executed using manually created rules).

**Table 2.** Compensations for transfer operations.

| **Transfer operation** | **Compensation** |
|---|---|
| Transfer of a word | 1.5 |
| Transfer of a NE | 0.4 |

After computing the final penalty value (the sum of all penalties imposed in step 1, reduced by compensations from step 2), the transfer score is arrived at using

the following formula:

$$score = 1 - \frac{p}{avgLength} \tag{1}$$

p - total penalty
avgLength - average number of words and Named Entities in the source and input sentences.

## 4 Evaluation of the transfer score metric

The transfer score metric was evaluated using both automatic and human-aided techniques. Automatic evaluation involved the comparison of the transfer score metric with other automatic machine translation quality metrics - BLEU ([4]) and METEOR ([9]). Human-aided evaluation, on the other hand, was aimed at finding out whether the transfer score metric can serve as a measure of usability of produced translations.

### 4.1 Test procedure

The JRC Corpus ([10]) was used for tests. The procedure of computing scores of test translations is presented in Figure 5. Following the preparation of the set of

**Algorithm 1: Test translation procedure**

```
select 10000 units from Polish-English JRC Corpus at random
for each selected unit u
    s := u.source
    r := u.target

    train EBMT sytem with JRC Corpus without unit u
    t := translation of sentence s by EBMT system
    transfer_score := score of translation of sentence s

    meteor_score := METEOR score for test translation t and reference r
    bleu_score := BLEU score for test translation t and reference r

    store the triple (transfer_score, meteor_score, bleu_score) in a file
```

**Fig. 5.** Algorithm for computing scores of test translations

scores, two correlations were analyzed separately: of the transfer and METEOR scores and of transfer and BLEU scores. To compute the correlation of two scores,

they were treated as random variables X and Y. The first step of finding their correlation was computing their covariance from the following formula:

$$cov(X,Y) = \sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) \tag{2}$$

Where:
$x_i, y_i$ - individual scores for translations
$\bar{x}, \bar{y}$ - average scores

The next step was computing standard deviations of the two variables from the formula:

$$\sigma(X) = \sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2} \tag{3}$$

The final step was computing the Pearson product-moment correlation coefficient ([11]) from the formula:

$$r_{XY} = \frac{cov(X,Y)}{\sigma(X)\sigma(Y)} \tag{4}$$

### 4.2 Comparison with METEOR

Figure 6 shows the chart of correlation between the transfer score and the METEOR score (it also shows the regression line). The computed Pearson product-moment correlation coefficient of the two measures in this case was **0.25**.

### 4.3 Comparison with BLEU

Figure 7 shows the chart of correlation between the transfer score and the BLEU score (as well as the regression line). The computed Pearson product-moment correlation coefficient of the two measures in this case was **0.32**.

### 4.4 Comparison with "human metric"

Apart from automatic evaluation, the transfer score metric was also evaluated by humans. Two Polish translators were asked to translate 20 Polish sentences from the JRC Corpus into English. For each sentence, they were given a suggestion for translation - a translation of the sentence performed by the EBMT system. The translators, however, were not provided with the translation scores for these translations. Instead, they were asked to judge the usability of suggestion using a 5 point scale (5 - no need to alter the suggestion, a good translation, 1 - suggestion to be ignored, need to translate the sentence from scratch). The score (marked as **p**) was then rescaled using the formula:

$$score = 0.25 * (p - 1) \tag{5}$$

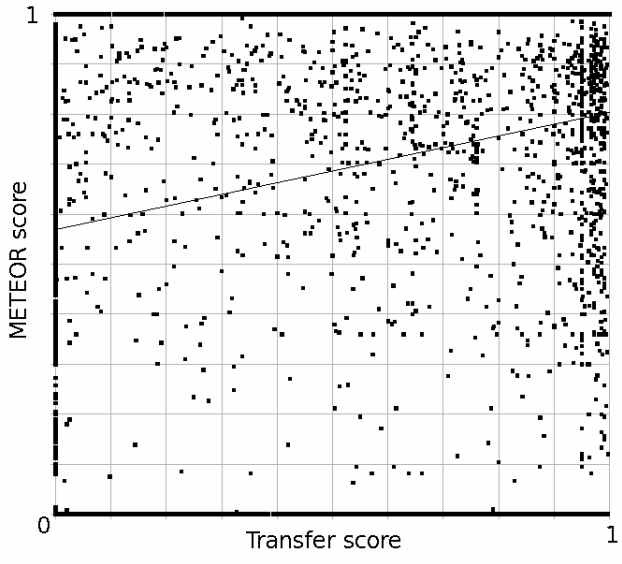


**Fig. 6.** The correlation between the transfer score and METEOR score

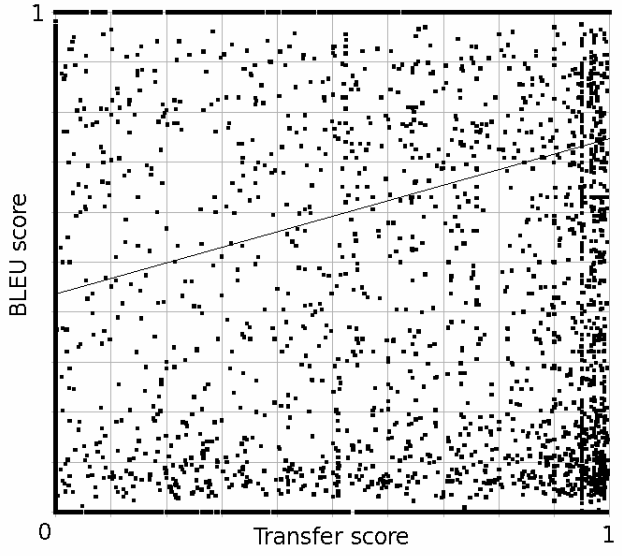


**Fig. 7.** The correlation between the transfer score and BLEU score

The final human score for each translation was computed as an average of the two people's scores.
The correlation of the transfer score and the human scores was arrived at in a

way similar to the previous evaluations. Figure 8 shows the correlation of the transfer score and the human scores. The computed Pearson product-moment

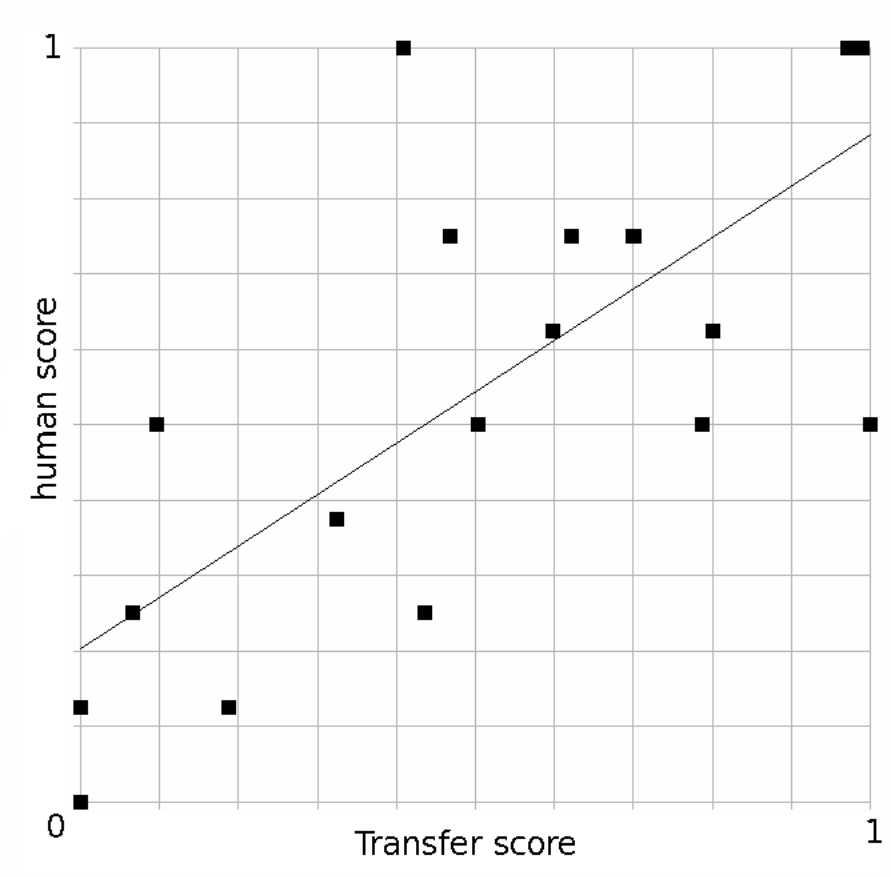


**Fig. 8.** The correlation between the transfer score and the human judgement

correlation coefficient for the transfer score metric and human judgment was **0.74**.

For comparison, the correlation of the human scores and the METEOR score was also computed in a similar way. Figure 9 shows the correlation of the METEOR score and the human scores. The computed Pearson product-moment correlation coefficient for the METEOR metric and human judgment was only **0.39**.

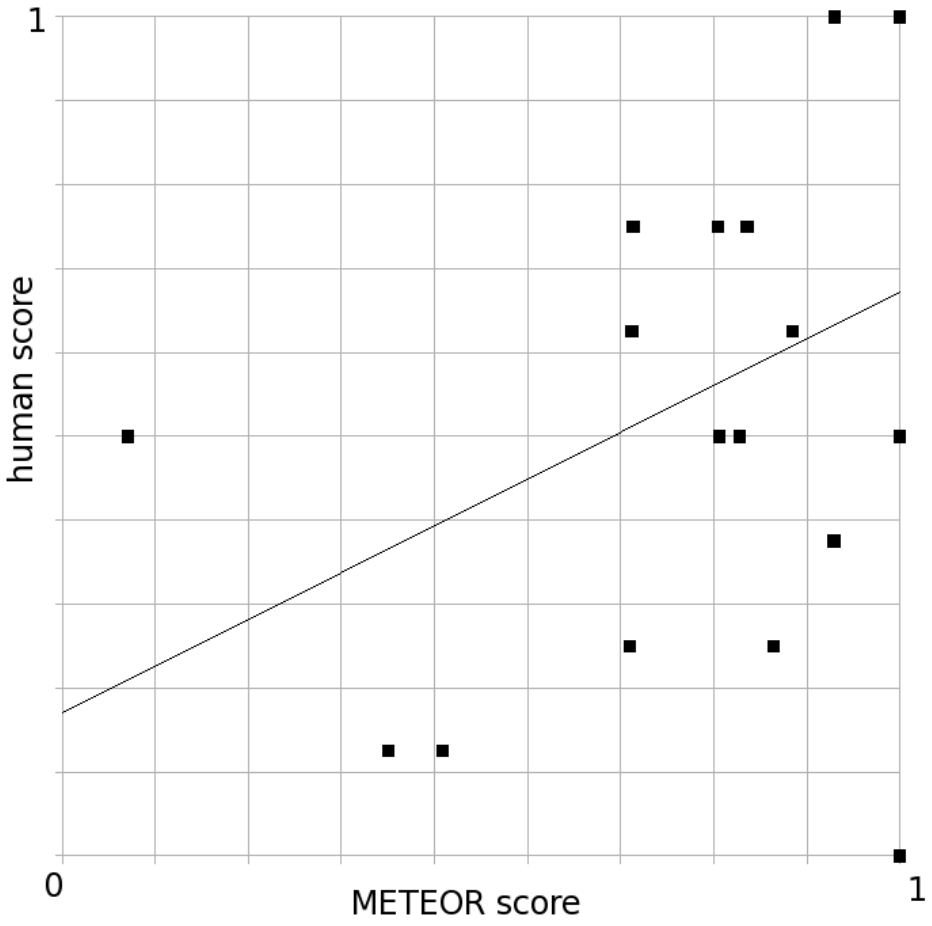


**Fig. 9.** The correlation between the METEOR score and the human judgement

## 5 Conclusions

The evaluation showed that the transfer score metric used in the described EBMT system is not much correlated with the METEOR and BLEU metric. It is, however, unlike METEOR, well correlated with human judgment of usability of translation. Therefore, the transfer score can be used in a CAT system to help the user (translator) decide, to what extent can the translation suggestion of EBMT systems be relied on.
Evaluation of EBMT systems should not be carried out using the same techniques as with other machine translation systems, especially when the EBMT system serves as a CAT system.

## References


1. Somers, H., Dandapat, S., Naskar, S.K.: A review of ebmt using proportional analogies. Proceedings of the 3rd International Workshop on Example-Based Machine Translation (2009)
2. Vandeghinste, V., Martens, S.: Top-down transfer in example-based mt. Proceedings of the 3rd International Workshop on Example-Based Machine Translation (2009)
3. Kurohashi, S.: Fully syntactic example-based machine translation. Proceedings of the 3rd International Workshop on Example-Based Machine Translation (2009)
4. Papineni, K., Roukos, S., Ward, T., Zhu, W.J.: Bleu: a method for automatic evaluation of machine translation. http://www1.cs.columbia.edu/nlp/sgd/bleu.pdf (2002)
5. Hongxu, H., Dan, D., Gang, Z., Hongkui, Y., Yang, L., Deyi, X.: An EBMT system based on word alignment. http://www.mt-archive.info/IWSLT-2004-Hou.pdf (2004)
6. Rapp, R., Vide, C.M.: Example-based machine translation using a dictionary of word pairs. http://www.mt-archive.info/LREC-2006-Rapp.pdf (2006)
7. Jassem, K., Marcińczuk, M.: Semi-supervised learning rule acquisition for Named Entity recognition and translation. unpublished (2008)
8. Gintrowicz, J.: Tłumaczenie automatyczne oparte na przykładach i jego rozszerzenia. Master thesis under the supervision of dr Krzysztof Jassem (2007)
9. Lavie, A., Agarwal, A., Denkowski, M.: The meteor metric for automatic evaluation of machine translation. http://www.cs.cmu.edu/∼alavie/METEOR/meteor-mtj-2009.pdf (2009)
10. Ralf, S., Pouliquen, B., Widiger, A., Ignat, C., Erjavec, T., Tufiş, D., Varga, D.: The jrc-acquis: A multilingual aligned parallel corpus with 20+ languages. Proceedings of the 5th International Conference on Language Resources and Evaluation (2006)
11. Stigler, M., S.: Francis galton's account of the invention of correlation. Statistical Science (1989)