Selecting Translation Strategies in MT using Automatic Named Entity Recognition

Bogdan Babych
Centre for Translation Studies
University of Leeds, UK
Department of Computer Science
University of Sheffield, UK
bogdan@comp.leeds.ac.uk

Anthony Hartley
Centre for Translation Studies
University of Leeds, UK
a.hartley@leeds.ac.uk

Abstract

We report on the results of an experiment aimed at enabling a machine translation system to select the appropriate strategy for dealing with words and phrases which have different translations depending on whether they are used as proper names or common nouns in the source text. We used the ANNIE named entity recognition system to identify named entities in the source text and pass them to MT systems in the form of "do-not-translate" lists. A consistent gain of about 20% in translation accuracy was achieved for all tested systems. The results suggest that successful translation strategy selection is dependent on accurate segmentation and disambiguation of the source text – aspects which could be significantly improved by named entity recognition. We further suggest an automatic method for distinguishing and lexical differences in MT output that could have applications in automated MT evaluation for morphologically rich languages.

1. Introduction

Language communities develop certain acceptable practices and norms for translating different types of concepts, expressions and texts from other languages and cultures. These practices are described as *translation methods*, *translation strategies* and *translation procedures*. (Vinay and Darbelnet, 1958, 1995). Translation methods relate to whole texts, while strategies and (finer-grained) procedures relate to sentences and smaller units (Newmark, 1988:81). The choice of a translation strategy often depends on the type of a translated unit. For example, for certain types of proper names the optimal translation strategy is *transference*, i.e., a "*do-not-translate*" or "*transliterate*" strategy, while the majority of common nouns are translated with other strategies: *literal translation, transposition, modulation,* etc. (Newmark, 1988: 81-88). This implies that recognising different types of units in the source text is a necessary condition for optimising the choice of translation strategy and, ultimately, for improving the quality of the target text.

The problem of selecting translation strategies for words that may be used as proper names or common nouns in the source language is related to a more general problem of word sense disambiguation (WSD) – one of the most serious problems for Machine Translation technology. Dealing with "proper *vs* common disambiguation" (PCD) often requires combining different knowledge sources, in a similar way to WSD (Stevenson and Wilks, 2001). But the cross-level nature of this problem also suggests that improvement in MT quality could be achieved through improving related aspects of the source-text analysis, such as Named Entity (NE) recognition (Babych and Hartley, 2003; Somers, 2003:524). For the purposes of this discussion, we assimilate proper nouns to NEs and investigate NE recognition as a possible solution to the PCD problem insofar as it might enable the selection of the correct strategy.

Accurate NE recognition is important for the general quality of MT for the following reasons:

1. The translation of the same token may be different depending on whether the token is a

common noun or part of an NE, e.g. in Russian if a common name is a part of an organization name, a "do-not-translate" or "transliterate" strategy should be used instead of a default translation strategy:

(1) **Original**: …the Los Angeles office of the *Hay Group*, a management consulting firm.
MT output[1]: …Лос-Анджелесский офис *Группы Сена*, управление консультантская фирма.
('… the Los Angeles office of the group of the hay [i.e., the grass, cut and dried for fodder], management consulting firm ')
Human translation: Лос-Анджелесский офис *Hay Group*, управленческой консультантской фирмы.

In this case NE recognition is directly linked to the PCD problem: we need to disambiguate between "common" and "NE" readings of the same string.

2. Failure to recognise NEs as single syntactic units or to determine their correct morpho-syntactic category in the source text may cause segmentation errors, which lead to the wrong morpho-syntactic structure in the target text, e.g.:

(2) **Original**: a *Big Board* spokesman couldn't comment on the talks.
MT output: *Большой* представитель *Правления* не мог комментировать переговоры.
('A big spokesman of the Board [management] couldn't comment on the talks').

In this case, NE recognition affects mainly morpho-syntactic segmentation, but individual words normally have correct translation strategies. However, a different morpho-syntactic context often requires the selection of a different translation strategy (either within or outside NEs), which may cause PCD errors in MT output, so there is an indirect link between morpho-syntactic disambiguation and PCD e.g.:

(3) **Original**: Moody's *Investors Service Inc*. placed the long-term debt under review.
MT output: *Инвесторы* Муди *Обслуживают компанию*, поместил долгосрочный долг под обзором.
('Investors of Moody serve the company, he placed the long-term debt under review').

Here the NE *Investors Service Inc*. is not treated as a single segment, which causes a combined morpho-syntactic and PCD error: the system translates the word *service* as a verb that means 'to serve' instead of using the correct "do-not-translate" strategy.

Thus NE recognition could be beneficial both for morpho-syntactic well-formedness and for correct PCD in MT output. In (Babych and Hartley, 2003) we addressed the first of these two problems. In this paper, we concentrate on the second problem and show how PCD can be improved using existing NE recognition modules.

Certain types of NEs, such as organisation names, appear to be a weak point even for some leading-edge MT systems, such as Systran and Reverso. At the same time, the problem of accurate NE recognition has been specifically addressed and benchmarked by the developers of information extraction (IE) systems. For example, the NE recognition module of the ANNIE IE system achieves a combined Precision & Recall score of 80-90% on news texts (Cunningham et al., 2002). Our suggestion is that combining this highly accurate NE recognition module with state-of-the-art MT systems would be beneficial for MT output, even if we do not change any of the other MT components.

The source code for commercial MT systems is not publicly available, so for our experiment we used one of the pre-processing tools of these systems – "do-not-translate" (DNT) lists. These lists were created from NE annotation produced by the ANNIE NE recognition module. For each of the three available MT systems we generated two different translations: a baseline translation and the DNT-processed translation. We made an approximate distinction between PCD and morpho-syntactic differences automatically using statistical frequency weights similar to tf.idf scores. We evaluated the improvement in PCD by manually annotating the PCD differences in the baseline and NE-processed MT output.

The remainder of the paper is organised as follows: in section 2 we discuss the rationale of our automated method for distinguishing lexical and morpho-syntactic differences in MT output. In section 3 we describe the linguistic resources and scoring procedure used in the experiment. In section 4 we present the PCD improvement achieved for three MT systems. Section 5 points out possible applications of the work to automatic MT evaluation. In section 6 we discuss conclusions and future work.

[1] The examples are taken from the output of MT systems that translated 30 texts of MUC-6 data, which was originally used for evaluating NE recognition.

2. Distinguishing lexical and morpho-syntactic differences in MT output

DNT-processing causes both morpho-syntactic and lexical differences in compared translations. In example (4) we annotate lexical (L) and morpho-syntactic (M) differences in the reference and DNT-processed translations. These differences are due to the fact that the company name "Eastern (Airlines)" received a correct morpho-syntactic category as a result of DNT-processing (Noun, not Adjective). Moreover, not translating this company name is the correct option for Russian target text.

(4) **Original:**
By proposing a meeting date, **Eastern** moved one step closer toward reopening current high-cost contract agreements

Baseline translation:
Предлагая дату встречи, **Восточный**-*(L)* перемещенный-*(M)* один шаг ближе к повторному открытию высокой стоимости-*(M)* потока-*(L)* заключают-*(L)* соглашения-*(M)*
('By proposing a meeting date, **Eastern** *(Adj.)* moved *(Participle)* one step closer toward reopening the high-cost$_{(ACC)}$ of a current *(Noun: 'the stream [of water, etc.]')* (they) conclude *(Verb)* agreements$_{(ACC)}$')

DNT-processed translation:
Предлагая дату встречи, **Eastern**-*(L)* переместил-*(M)* один шаг ближе к повторному открытию текущих-*(L)* соглашений-*(M)* контракта-*(L)* с высокой стоимостью-*(M)*
('By proposing a meeting date, **Eastern** *(Noun)* moved *(Verb)* one step closer toward reopening of current *(Adj.)* agreements$_{(GEN)}$ of a contract *(Noun)* with high cost$_{(INST)}$')

	Original	Baseline	DNT-proc.
L	Eastern	Восточный ('Eastern$_{(ADJ)}$')	Eastern (not translated)
L	Current	потока (stream$_{(NOUN)}$')	текущих ('current$_{(ADJ)}$')
L	Contract	заключают ('conclude$_{(VERB)}$')	контракта ('contract$_{(NOUN)}$')
M	Moved	перемещенный (PARTICIPLE)	переместил$_{(VERB)}$
M	Cost	стоимости$_{(GEN)}$	стоимостью$_{(INST)}$
M	Agreements	соглашения$_{(ACC)}$	соглашений$_{(GEN)}$

Table 1. Examples of translation differences

In this example, all six variants in the DNT-processed translation are better than their counterparts in the baseline translation.

Note that a correct PCD choice for *lexical* differences is determined by the senses of the words in the source text, and there is no way of correctly using lexical items from the baseline translation as alternative translations. In contrast, the source text does not require particular values of *morpho-syntactic* categories in the target text. These values are determined by the rules of the target language and by the morpho-syntactic structure of a sentence, chosen by a translator. In many cases these values can be subject to greater variation then the lexical choices. For example, there is a legitimate way of using the last two words in the Table 1 in the genitive and accusative case, as in the baseline translation shown in example (5), if these values are required by their morpho-syntactic position:

(5) Предлагая дату встречи, Eastern переместился на один шаг ближе к тому, чтобы повторно открыть текущие контрактные соглашения$_{(ACC)}$ высокой стоимости$_{(GEN)}$.
('By proposing a meeting date, Eastern moved one step closer toward that [situation], to reopen current agreements$_{(ACC)}$ of high cost$_{(GEN)}$)

A rough distinction between morpho-syntactic and lexical differences in the compared output texts can be drawn automatically using term frequency weights proposed in (Babych, Hartley, Atwell, 2003) for evaluating MT for Information Extraction purposes. These weights (S-scores) are similar to tf.idf scores: they describe the relative salience of terms in a particular text. They were found to make an accurate distinction between content and function words. With a varying degree of accuracy (depending on how analytic the grammar of a given language is) this distinction also separates lexical and morpho-syntactic differences in compared texts. For Russian (which has a not highly analytic grammar) it achieves 88.4% Precision for lexical items, while for French the Precision is 98%.

The S-scores are computed for each word in each text using the following formula:

$$S(i,j) = \log \frac{\left(P_{doc(i,j)} - P_{corp-doc(i)}\right) \times (N - df_{(i)}) / N}{P_{corp(i)}}$$

where:

- $P_{doc(i,j)}$ is the relative frequency of the word in the text; ("Relative frequency" is the number of tokens of this word-type divided by the total number of tokens).
- $P_{corp\text{-}doc(i)}$ is the relative frequency of the same word in the rest of the corpus, without this text;
- $P_{corp(i)}$ is the relative frequency of the word in the whole corpus, including this particular text.
- df_i is the number of documents in the corpus where the word w_i occurs (the *document frequency*);
- N is the total number of documents in the corpus;

We computed S-scores for words with:

$(P_{doc(i,j)} - P_{corp\text{-}doc(i)}) > 0$; $AbsFrq_i > 1$, where $AbsFrq_i$ is the number of occurrences of the word w_i in the corpus.

Table 2 illustrates the ranking of words according to their S-score for one of the English texts from MUC6 NE corpus, for which $tf_{i,j} > 1$ ($tf_{i,j}$ is the number of occurrences of the word w_i in the document d_j).

r	S	word	r	S	Word
1	2,918	OPEC	8	1,844	total
1	2,918	Emirates	8	1,844	report
1	2,918	barrels	9	1,692	current
1	2,918	oil	10	1,593	price
1	2,918	quota	10	1,593	news
1	2,918	Subroto	11	1,470	recent
1	2,918	world	12	1,270	month
1	2,918	cartel	13	1,161	officials
1	2,918	war	14	0,972	because
1	2,918	ruler	15	0,805	million
1	2,918	petroleum	16	0,781	yesterday
1	2,918	markets	17	0,651	that
1	2,918	gestures	18	0,621	also
1	2,918	estimates	19	0,527	much
1	2,918	conciliatory	20	0,331	but
1	2,918	Zayed	21	0,291	over
1	2,918	UAE	22	0,007	from
1	2,918	Szabo	23	-0,079	there
1	2,918	Sheik	24	-0,126	after
1	2,918	Saudi	25	-0,233	their
1	2,918	Petroleum	26	-0,244	new
1	2,918	Dhabi	27	-0,284	had
1	2,918	Arabia	28	-0,411	as
1	2,918	Abu	29	-1,225	talks
2	2,719	output	30	-1,388	been
3	2,449	others	31	-1,594	at
3	2,449	manager	33	-1,844	on
3	2,449	government	34	-2,214	its
3	2,449	dropped	35	-3,411	for
3	2,449	declines	36	-3,707	with
3	2,449	agency	38	-4,238	the
4	2,375	day	39	-4,319	by
5	2,305	production	40	-4,458	Mr
6	2,096	well	41	-5,323	the
6	2,096	demand	42	–	a
7	1,880	concern	42	–	of

Table 2. Ranking of words by the S-score

We established by experiment that a reasonable threshold for distinguishing content words and functional words is:

S-score = 1

This threshold gives good results for text in all analysed languages: English, French and Russian. Our assumption implies that for comparing lexical differences in two variants of translation we need to compare for each text sets of words with an S-score above the threshold.

Accordingly, all words that were different in each set were automatically highlighted in their respective texts and presented for manual scoring. In the examples of MT in the following sections, words with $tf_{i,j} > 1$ are bold, words with $tf_{i,j} = 1$ are bold and italic. In the original English sentences, the NEs used for the DNT lists are highlighted in bold.

3. Resources and scoring method

For our experiment we used the following linguistic resources: 30 texts (news articles) which were processed with the NE recognition module of the GATE-1 IE system in the DARPA MUC6 competition. The results of manual NE annotation were also available, but GATE NE recognition is sufficiently accurate for these texts (Recall – 84%, Precision – 94 %, Precision and Recall – 89.06% (Gaizauskas et al, 1995)) that errors in the GATE output will not have had a major impact on our results.

Table 3 summarises the statistical parameters of the corpus analysed. The corpus is rich in NEs, so the effect of NE recognition on PCD could be accurately measured for the MT systems.

Number of:	*For the corpus*	*Av. per doc.*	*Av. per para.*	*Av. per sent.*
Paragraphs	283	9.4	–	–
Sentences	565	18.8	2.0	–
Word occurrences	11975	399.2	42.3	21.2
Different words	3944	235.7	36.3	19.7
NE occurrences keys/ ***GATE***	544/ **510**	18.1/ **17.0**	1.9/ **1.8**	1.0/ **0.9**
Different NEs: keys/ ***GATE***	201/ **174**	7.6/ **6.7**	1.5/ **1.4**	0.9/ **0.8**

Table 3: Statistical parameters of the corpus

DNT lists were automatically generated from GATE annotations and the texts were translated with three commercial MT systems:

- English-Russian 'ProMT 98' v4.0, released in 1998
- English-French 'ProMT', (*Reverso*) v5.01, released in 2001
- English-French 'Systran Professional Premium' v3.0b, released in 2000

Two translations were generated by each MT system:

- **a baseline translation** without a DNT list
- **a DNT-processed translation** with the automatically created DNT list of organisation names

The baseline and the DNT-processed translation were automatically compared using the method presented in Section 2. Lexical differences were highlighted and scored according to the following criterion:

+1 – PCD is correct in the DNT-processed translation and is wrong in the baseline translation

0 – PCD in both translations is equally (not) correct

–1 – PCD is wrong in the DNT-processed translation, or DNT-processing is not acceptable translation strategy for the NE; PCD is correct in the baseline translation

Further examples illustrate these scores:

+1	**Original:** A week earlier, **Eastern** sued the Machinist and pilot unions
	Baseline translation: Неделей ранее, **Восточный~+1** ***преследуемый~+1*** перед Машинистом и экспериментальными союзами. ('A week earlier, Eastern$_{(ADJ)}$ (was) chased$_{(Participle)}$ before the Machinist and experimental unions')
	DNT-processed translation: Неделей ранее, **Eastern~+1** предъявил ***иск~+1*** Машинисту и экспериментальным союзам ('A week earlier, **Eastern**$_{(NOUN)}$ brought **suit**$_{(NOUN)}$ against the Machinist and experimental unions')

+0	**Original:** About 6,000 salaried workers are currently represented by the **United Auto Workers** union.
	Baseline translation: Приблизительно 6,000 оплачиваемых рабочих в настоящее время представлены Объединенным союзом Работников автомобильной промышленности***~0***. ('About 6,000 salaried workers are currently represented by the United union of Workers of automobile industry.')
	DNT-processed translation: Приблизительно 6,000 оплачиваемых рабочих в настоящее время представлены союзом United ***Auto~0*** Workers. ('About 6,000 salaried workers are currently represented by the union "United Auto Workers".')

–1	**Original:** **Treasury** Secretary James Baker held a 7 1/2-hour negotiating session with top Canadian officials.
	Baseline translation: **Министр~–1** финансов Джеймс Бакер проводил 7 1/2-часовых сессии ведения переговоров с высшими Канадскими должностными лицами ('The minister of finances James Baker held a 7 1/2-hour negotiating session with top Canadian officials..') – correct translation equivalent chosen
	DNT-processed translation: Секретарь**~–1** Treasury, Джеймс Бакер проводил 7 1/2-часовых сессии ведения переговоров с высшими Канадскими должностными лицами ('Secretary of "Treasury" James Baker held a 7 1/2-hour negotiating session with top Canadian officials.') – incorrect translation equivalent
	Original: The **Labour Department** has collected the statistics.
	Baseline translation: **Министерство~–1** труда**~–1**, собрало статистику. ('The Ministry of Labour has collected the statistics.')
	DNT-processed translation: **Labor~–1 Department~–1**, собрало статистику. ('The **Labor Department** has collected the statistics.') – unacceptable translation strategy

All differences highlighted in the whole MUC-6 NE corpus were manually annotated for each of the MT systems under consideration. Cases of morpho-syntactic differences were also annotated and excluded from the scored set of differences. The number of annotated differences is presented in Table 4:

	ProMT 1998 E-R	*ProMT 2001 E-F*	*Systran 2000 E-F*
Highlighted differences;	528	161	176
Including:			
diff.	61	3	2
scored lexical diff./Precision	467 (88.4%)	158 (98.1%)	174 (98.9%)

Table 4

The larger number of differences and the lower Precision for the Russian system can be attributed to the largely synthetic morphology of Russian.

The overall score for improvement / decline in PCD for each MT system was calculated as a sum of all scores of lexical differences divided by the number of lexical differences for the particular system.

4. Results of the experiment for PCD

The set-up of this experiment gives a reasonable estimate of the influence of NE recognition on MT quality, and suggests that if improvement in MT can be achieved via pre-processing tools, then we can expect even greater improvement when an NE recognition module is properly integrated into MT systems (e.g., types of NEs requiring non-transference translation strategies are also distinguished). The improvement achieved for the MT systems under consideration was around 20%.

The results of manual annotation are summarised in Table 5:

	ProMT 1998 E-R		*ProMT 2001 E-F*		*Systran 2000 E-F*	
Mark	*N*	*Score*	*N*	*Score*	*N*	*Score*
+1*	154	+154	62	+62	77	+77
0*	239	0	66	0	61	0
–1*	74	–74	30	–30	36	–36
∑	467	+ **80**	158	+ **32**	174	+ **41**
Gain		**+17.1%**		**+20.2%**		**+23.6%**

Table 5 Scoring results

All systems showed consistent improvement in PCD tasks after NE recognition. The results indicate that systematic NE recognition has great potential for improving the quality of MT, and that successful PCD depends on appropriate analysis of other aspects in the source text, such as determining correct values for morphological categories and correct syntactic segmentation. These aspects could be substantially improved via NE recognition.

However, finding appropriate segmentation and morpho-syntactic disambiguation is a necessary but not a sufficient condition for achieving improvement in MT: most cases of decline in MT quality after DNT-processing are due to the lack of flexibility in determining the optimal translation strategy for NEs. In our experiment, the overall improvement in the quality of PCD is due to the fact that the *transference* ("*do-not-translate*") strategy is optimal, or it is an acceptable translation strategy for the majority of NE that occurred in our corpus (Newmark, 1982). But many NEs might need to be translated by specific translation equivalents that are normally recognised by the state-of-the-art MT systems. This is especially important for names of well-known organisations, such as 'The Treasury', 'The Army', 'The Navy' 'Labour', which are often part of more complex NEs: 'The Treasury Secretary', 'The Labour Government', 'The Army Chief' – in all these cases a "do-not-translate" strategy could cause a serious decline in MT quality.

Our analysis suggests that targeting specific needs of MT could be a way of improving MT quality with IE technology: the NE recognition stage could meet the needs of MT systems by distinguishing different classes of NEs which require different translation strategies. Appropriate annotation of these NEs in the source text could then guide the MT system at the transfer stage.

5. Conclusions and future work

We have characterised the potential improvement in PCD for MT systems achievable with accurate NE recognition. The results indicate that PCD is very sensitive to those aspects of MT quality which can be improved with NE recognition: finding appropriate morpho-syntactic categories and correct segmentation for NEs often influences the correctness of the general analysis of the source sentence. But some aspects of PCD cannot be improved with existing NE recognition and need to be addressed by the IE and MT communities jointly. NE recognition modules can be extended to distinguish between types of NEs that require different translation strategies; and MT systems can be adapted to deal more flexibly with user input, by using NE annotation designed specifically for MT purposes.

The proposed method of making a rough automatic distinction between lexical and morpho-syntactic differences allowed us to annotate important features in a relatively large corpus within a reasonable amount of time. We suggest that this method could have applications in other domains of NLP, in particular – in automated MT evaluation and in automatic alignment of parallel texts.

5.1 Application to automatic MT evaluation

Current automatic evaluation methods, such as BLEU (Papineni et al., 2001), do not make a distinction between lexical and morpho-syntactic differences, but distinguishing them and controlling the quality of MT on several separate levels might be useful to for the evaluation of MT systems under development (especially for target languages with a rich morphology, where these two types of differences clearly characterise different aspects of quality).

Another important problem for further research is establishing whether different degrees of legitimate variation in translation are allowed for items with different tf.idf and S-scores. One of the most serious problems for the BLEU method is related to legitimate variability in the reference translation. In order not to penalise acceptable MT that is different from human translation, the metric uses several reference translations of the same text. These resources can be expensive to create. However, if terms with different significance scores show different levels

of legitimate variation, then the metric could rely on potentially more stable terms, so fewer reference texts would be needed to produce consistent evaluation scores for MT systems.

Yet another problem for the BLEU metric is high data scarcity of N-grams in languages with complex synthetic morphology, such as Slavonic languages. In order to achieve evaluation scores comparable with scores for English or other analytical languages, we need to use much larger reference corpora of human translations. An alternative solution to this problem could be to make automatically a rough distinction between lexical and morphological differences and to concentrate on the lexical differences that are expected to be less sparse across human translations and MT output.

5.2 Application to automatic alignment of parallel texts

An analysis of S-scores (Section 2) of lexical differences in the compared translations also gives interesting results. It can be noted that words which are translations of the same word in the DNT-processed and the baseline target texts have very close scores. Ranked lists of differences for Russian MT are presented in Table 6:

DNT-processed translation	**Baseline translation**
1:NBC:3.939817	1:ЭН-БИ-СИ:3.906120
1:***Техники***:3.416626 technicians$_{(NOM.PLUR)}$	1:***Техников***:3.382496 (of) technicians$_{(GEN.PLUR)}$
1:***Electric***:3.416626	1:***Электрическая***:3.382496 electric$_{(NOM.SING.FEM)}$
1:***Broadcast***:3.416626	1:***Радиопередачи***:3.382496 of broadcast$_{(GEN.SING)}$
2:***Служащие***:2.959119 employees$_{(NOM.PLUR)}$	2:***Служащих***:2.924432 of employees$_{(GEN.PLUR)}$
2:***General***:2.959119	2:***Общая***:2.924432 general$_{(NOM.PLUR.FEM)}$
3:***Association***:1.886203	3:***Ассоциации***:2.303370 of association$_{(GEN.SING)}$

Table 6 Scores for corresponding words

The match between S-scores is closer for words with a unique translation, which implies that they have similar distribution in the text and in the corpus.

Another interesting property of the statistical significance measure is that different word forms which are translations of the same word (e.g., an English NE) often have very close S-scores, which are also close to the score of the original word. For example, S-scores for the first word in the NE "Pan Am" and for three morphological variants of its wrong translation into Russian are presented in Table 7. All are variants of the lexeme "кастрюля" – 'saucepan', and also have different frequencies in the text. This effect is also the strongest for words which have a unique translation in the corpus.

DNT-NE / S-score	*Abs. frq. in DNT text / in the rest of corpus*	*Baseline transl. of NE*	*Abs. frq. in baseline text / in the rest of corp.*
Pan 3.087052	14 / 0	Кастрюля$_{(NOM)}$ 3.112597	8 / 0
		Кастрюлю$_{(ACC)}$ 3.112597	2 / 0
		Кастрюли$_{(GEN)}$ 3.112597	2 / 0

Table 7 Scoring results

This property of the S-score may be useful in MT evaluation for highly inflected languages.

Future work in this direction will involve measuring the accuracy of the suggested method of distinguishing morpho-syntactic and lexical differences in MT output for typologically different languages and evaluating the degree of legitimate variation in translation at different levels of the significance scores.

References

Babych, B., A. Hartley and E. Atwell. 2003. Statistical Modelling of MT output corpora for Information Extraction. In: *Proceedings of the Corpus Linguistics 2003 conference, edited by Dawn Archer, Paul Rayson, Andrew Wilson and Tony McEnery*. Lancaster University (UK), 28 - 31 March 2003. Pp. 62-70.

Babych, B. and A. Hartley. 2003. Improving Machine Translation Quality with Automatic Named Entity Recognition. In *Proceedings of the 7th International EAMT workshop on MT and other language technology tools. Improving MT throught other language technology tools. Recourses and tools for building MT.* Budapest, Hungary. p. 1–8.

Cunningham, H., D. Maynard, K. Bontcheva, V. Tablan. 2002. GATE: A Framework and Graphical Development Environment for robust NLP Tools and Applications. *Proceedings of the 40th Anniversary Meeting of the Association for Computational Linguistics (ACL'02).* Philadelphia, July 2002.

Gaizauskas, R., T. Wakao, K. Humphreys, H. Cunningham, Y. Wilks. 1995. University of Sheffield: Description of the LaSIE system as used for MUC-6. *Proceedings of the 6th Message Understanding Conference (MUC-6).* Morgan Kaufmann, pp. 207-220.

Newmark, P. 1982. Approaches to translation. Pergamon Press, Oxford, NY.

Newmark, P. 1988. A textbook of translation. Longman, London, NY.

Papineni, K., S. Roukos, T. Ward, and W-J Zhu. 2001. Bleu: a method for automatic evaluation of machine translation. IBM research report RC22176 (W0109-022) September 17, 2001.

Somers, H. 2003. Machine Translation: latest developments. In: *The Oxford handbook on Computational Linguistics. Ed. By Ruslan Mitkov.* Oxford University Press, Oxford, NY. – Pp. 512-528.

Stevenson, M. and Y. Wilks. 2001. The integration of knowledge sources in word sence disambiguation. Computational Linguistics 27(3):321-349.

Vinay, J.P. and J.Darbelnet. 1958. Sylistique comparée du français et de l'anglais: Methode de traduction. Didier, Paris.

Vinay, J.P. and J.Darbelnet. 1995. Comparative stylistics of French and English : a methodology for translation / translated and edited by Juan C. Sager, M.-J. Hamel. J. Benjamins Pub., Amsterdam, Philadelphia.