Improving Machine Translation Quality with 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

Named entities create serious problems for state-of-the-art commercial machine translation (MT) systems and often cause translation failures beyond the local context, affecting both the overall morphosyntactic well-formedness of sentences and word sense disambiguation in the source text. We report on the results of an experiment in which MT input was processed using output from the named entity recognition module of Sheffield's GATE information extraction (IE) system. The gain in MT quality indicates that specific components of IE technology could boost the performance of current MT systems.

1. Introduction

Correct identification of named entities (NEs) is an important problem for machine translation (MT) research and for the development of commercial MT systems. In the first place, translation of proper names often requires different approaches and methods than translation of other types of words (Newmark, 1982: 70-83). Mistakenly translating NEs as common nouns often leads to incomprehensibility or necessitates extensive post-editing. In many cases failure to correctly identify NEs has an effect not only on a local and immediate context, but also on the global syntactic and lexical structure of the translation, since proper segmentation of a source text might be seriously distorted.

However, the developers of commercial MT systems often pay insufficient attention to correct automatic identification of certain types of NE, e.g., organisation names. This is due partly to the greater complexity of this problem (the set of proper names is open and highly dynamic), and partly to the lack of time and other development resources.

On the other hand, the problem of correct identification of NE is specifically addressed and benchmarked by the developers of Information Extraction (IE) systems, such as the GATE system, created at the University of Sheffield and distributed under GPL (Cunningham et al., 1996, 2002). The quality of automatic NE identification has been evaluated at several message-understanding conferences (MUC) sponsored by DARPA. Accuracy scores for leading systems are relatively high (in comparison to other IE tasks, such as co-reference resolution, template element filling or scenario template filling). The default settings of NE recognition module of the GATE system produces between 80-90% Precision & Recall on news texts (Cunningham et al., 2002).

In this paper we describe the effect of using the GATE NE recognition module as a pre-processor for commercial state-of-the-art MT

systems. The idea of our experiment is that high-quality automatic NE recognition, produced by GATE, could be used to create do-not-translate (DNT) lists of organisation names, a specific type of NE which in human translation practice is often left untranslated. (Newmark, 1982: 70-83).

In our experiment we systematically analysed the effect of incorrect NE recognition on the surrounding lexical and morphosyntactic context in MT output. We tried to establish how far NE recognition (specifically recognition of organisation names) influences grammatical well-formedness and word sense choices in the context of NEs. We compared the baseline translations (produced without NE DNT-processing) with translations produced using DNT lists (created with the GATE-1 NE recognition system), by systematically scoring cases of improvement and decline in lexical and morphosyntactic well-formedness. Texts with NE DNT-processing showed consistent improvement for all systems in comparison with baseline translations. The improvement was not lower than 20%.

This indicates that combining present-day MT systems with specific IE modules (where certain NLP problems are treated systematically) has beneficial effect on the overall MT quality.

2. Problems of NEs for MT

NEs usually require different approaches to translation than do other types of words. For example, foreign person names in Russian should be transcribed and written in Cyrillic; names that coincide with common nouns should not be looked up in the general dictionary. In some cases NEs (mostly organisation names) are not translated and preserve Roman orthography within Russian Cyrillic text. For example, in a 1000-word selection of 4 articles about the international economy on the Russian BBC World Service site, Roman-script NEs within the Cyrillic text covered 6% of the selection. The following NEs were neither translated, nor transliterated into Cyrillic: 'Nestle' (9 occurrences), 'AOL' (8); 'Buffalo Grill' (7); 'Burger King' (7); 'Diageo' (7); 'Schweisfurth (Group)' (2). In general, the practice not to translate organisation names is very common for translations into Slavic languages.

Mistakes related to the failure to distinguish between common nouns and proper nouns in MT can be very serious. For example, in our experiments an MT system translated the person name *Ray* as *Луч* ('beam of light'). Translating parts of compound NEs is also detrimental to MT quality, since it often involves incorrect segmentation of NEs: *American Telephone and Telegraph Corp.* was translated as *Американский Телефон и Компания Телеграфа* ('an American telephone and a company of a telegraph'). Yet another problem for MT systems is that failure to recognise NEs often has a negative effect on well-formedness of morphosyntactic and lexical context beyond the NEs themselves. Certain morphological features of neighbouring and syntactically related words, word order, a choice of word senses in MT output could be distorted if a NE is not correctly recognised. For example, an English phrase (1) was translated into Russian as (2):

(1) **Original:** Eastern Airlines executives notified union leaders ...

(2) **MT output:** *Восточные исполнители Авиалиний уведомили профсоюзных руководителей ...*
('Oriental executives of the Airlines notified ...')

This happened because the failure to identify *Eastern Airlines* as a NE led to incorrect syntactic segmentation of the sentence.

However, current MT systems allow the processing of MT input with DNT lists. Making a DNT of organisation names from the text in most cases improves not only the acceptability of NE translation, but also the overall well-formedness of the morphosyntactic and lexical context. For example, after the string *Eastern Airlines* was entered into a DNT list for the English-Russian MT system, the translation of (1) was morphologically and syntactically well-formed:

(3) DNT-processed MT output: *Исполнители Eastern Airlines уведомили профсоюзных руководителей ...*

Creating DNT lists manually requires much effort from the user of an MT system. However, the high accuracy in NE tagging of current IE systems, including GATE, means that DNT lists for MT can be created automatically.

The performance results reported here are based entirely on automatically created DNT lists used to process NEs.

3. Description of the experiment

In order to measure the effect of NE recognition on MT quality, we took 30 texts (news articles) from the DARPA MUC-6 evaluation set. These texts were selected because they are relatively rich in NEs, and because clean NE annotation is available for them. We used the following linguistic resources of the Sheffield NLP group:
- DARPA 'keys' - texts manually annotated with NEs;
- GATE 'responses' - the output of the automatic NE annotation of the GATE-1 system, which participated in MUC-6.

Table 1 summarises statistical parameters of this corpus. The table indicates how frequently NEs (organisation names) occur and shows that GATE 'response' figures are very close to the DARPA "key" figures.

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 1 : Statistical parameters of the corpus

The density of NEs in the DARPA corpus is also characterised by Table 2:

	Manual keys	GATE
Paragraphs with NEs	228 (80.6%)	**218 (77.0%)**
Sentences with NEs	329 (58.2%)	**315 (55.8%)**

Table 2: NE density in the corpus

The accuracy of GATE-1 in the NE recognition task at MUC-6 (Recall - 84%, Precision - 94%, Precision & Recall - 89.06 % (Gaizauskas et al., 1995)) is such that we used the GATE output for our MT experiment, rather than the cleaner manually annotated data. Moreover, the advantage of using automatic NE recognition is that the results of the experiment should be consistent with the results for other corpora on which the NE recognition task has been performed.

Having automatically generated DNT lists of organisation names from GATE 'response' annotation, we translated the texts using three commercial MT systems:
- English-Russian 'ProMT 98' v4.0, released in 1998 (Softissimo)
- English-French 'ProMT', *(Reverso)* v5.01, released in 2001 (Softissimo)
  English-French 'Systran Professional Premium' v3.0b, released in 2000 (Systran)

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 translations were then compared with DNT-processed translations, with respect to the morphosyntactic well-formedness of the context surrounding the NEs.

3.1. Segmentation

To speed-up the process of finding contextual differences, we developed automatic tools, which allowed us to make a formal distinction between NE-internal and NE-external issues in MT. Whereas Al-Onaizan and Knight (2002) focus on the former issue, our primary interest is in NE-external differences in context caused by improved NE recognition after DNT-processing. Thus, we automatically selected paragraphs with contextual differences and highlighted different strings in these paragraphs.

The example below illustrates the output of these annotation tools:
  Different strings found in two translations are indicated by '—>'
  'ORI' indicates the original English string in the DARPA corpus;
  'TWS' (baseline translation) indicates a String Translated Without the do-not-translate list;

'TDS' (DNT-processed translation) indicates a String Translated with Do-not-translate list.

—>40;TDSnotInTWS: 40# Отдельно, в его регистрации
—>40;TDSnotInTWS: раскрыл детали его планов финансирования приобретения

40;ORI=40#<s> Separately, in its <ENAMEX TYPE="ORGANIZATION">SEC</ENAMEX> filing, <ENAMEX TYPE="ORGANIZATION">USAir</ENAMEX> disclosed details of its plans for financing the <ENAMEX TYPE="ORGANIZATION">Piedmont</ENAMEX> acquisition.

40;TWS= 40# Отдельно, в ее регистрации СЕКУНДЫ, USAir раскрытые детали ее планов финансирования Предгорного приобретения.

40;TDS= 40# **Отдельно, в его регистрации** SEC, USAir **раскрыл детали его планов финансирования приобретения** Piedmont.

Since the amount of manual annotation was relatively small, no complex alignment for the two translated texts was implemented. Instead, we implemented a simple segmentation algorithm for paragraphs, using NE annotation in the corpus.

The segmentation was done in two stages. First, tagged NEs from the 'ORI' paragraph were identified and searched for in the 'TDS' paragraph. Then they were used as separators for the TDS: parts of the TDS between (untranslated) NEs were identified and searched for in the 'TWS' paragraph. If any sub-string was not found in TWS, it was printed and also highlighted in bold in TDS. This shows that strings in the context of the NE are different in the DNT-processed translation and in the baseline translation. This difference was then manually scored.

3.2. Scoring

Contextual differences between the baseline translation and the DNT-processed translation were manually scored using the scale in Table 3.

The terms 'well-formed' and 'not well-formed' refer to the local morphosyntactic or lexical context within a segment where differences occur. It remains possible that well-

formed structures require post-editing at a higher level in the translated text.

The term 'features' refers to morphosyntactic or lexical features of certain words in the context of the NE. By 'more correct', we mean that the features considered in the context are correct, but the corresponding features in the compared text are wrong.

Score	Baseline translation	DNT-processed translation
+ 1	not well-formed	well-formed
+ 0.5	not well-formed;	not well-formed; some features are more correct
= 0	equally (not) well-formed	
− 0.5	not well-formed; some features are more correct	not well-formed
− 1	well-formed	not well-formed

Table 3: Scoring scheme

Here are some example strings to illustrate each score:

+1	**Original:** (It) represents 4,400 *Western Union* **employees** around the country. **Baseline translation:** (Он) представляет 4,400 **Западных служащих Союза** по всей стране. ('It represents 4,400 **Western employees of the Union** around the country') **DNT-processed translation:** (Он) представляет 4,400 **служащих Western Union** по всей стране. ('(It) represents 4,400 **employees of Western Union** around the country')
+0.5	**Original:** Western Union Corp. **said its subsidiary,** Western Union Telegraph Co.... **Baseline translation:** Западная Корпорация Союза **сказала ее вспомогательную,** Западную Компанию Телеграфа Союза... ('Western Corporation of a Union said **its auxiliary (case.acc.),** Western Company of Telegraph of a Union ...') **DNT-processed translation:** Western Union Corp. **Сказанный его филиал,** Western Union Telegraph Co. ... ('Western Union Corp. **Its branch (case.nom) is said,** Western Union Telegraph Co....')

=0   **Original:**
American Airlines **Calls** for Mediation
Baseline translation:
Американские Авиалинии **Призывают** К посредничеству
(American Airlines **Call(num.plur.)** for Mediation)
DNT-processed translation:
American Airlines **Призывает** К посредничеству
(American Airlines **Calls(num.sing.)** for Mediation)

-0.5   **Original:**
USAir said **that** William R. Howard, chairman and chief executive *of Piedmont,* will be elected president of *USAir*
Baseline translation:
USAir сказал **тот** Уильям Р. Говард, председатель и руководитель Предгорных, будут избраны президентом USAIR
USAir said **that (particular)** (demonstr.pron,nom.) William R. Howard, chairman and chief executive of piedmont people, will be elected president of *USAir*
DNT-processed translation:
USAir сказал того Уильяма Ра. Говард, председатель и руководитель Piedmont, будут избраны президентом USAir
USAir said **of that (particular)** (demonstr.pron,gen.) William Ra. Howard, chairman and chief executive *of Piedmont,* will be elected president of *USAir*

-1   **Original:**
to discuss the benefits **of combining** *TWA* and *USAir*
Baseline translation:
чтобы обсудить выгоды от **объединения** TWA и USAIR
('to discuss the benefits **of the merge (noun) (of)** *TWA* and *USAir'*)
DNT-processed translation:
чтобы обсудить выгоды от **объединяющегося** TWA и USAir
('to discuss the benefits of the **combining (participle, sing.)** *TWA* and **(of)** *USAir'*)

For each MT system, we scored 50 strings showing differences. Table 4 summarises the number of paragraphs with contextual differences between the baseline and DNT-processed translations.

The figures in row 2 - *Paragraphs with contextual differences* — show to what extent DNT-processing affects the NE context for each system, showing also the percentage of these paragraphs in relation to the corresponding figure in row 1. Row 3 represents the percentage of manually scored paragraphs in relation to the

figure in row 2. These figures show the likely reliability of the results for manual scoring presented in the next section.

Number of:	*Original – GATE*	*MT E-R ProMT*	*MT E-F ProMT*	*MT E-F Systran*
Paras. with NE	218	225	225	239
Paras. with contextual differences		139 (61.8%)	132 (58.7%)	207 (86.6%)
Paras. manually scored		31 (22.3%)	28 (21.2%)	30 (14.5%)
Strings with differences		211	212	411
Strings scored		50 (23.7%)	50 (23.6%)	50 (12.2%)
Diff. strings per text		7.0	7.0	13.7
Diff. paras. per text		4.6	4.4	6.9

Table 4: Paragraphs with contextual differences

Note that in row 1 there is a mismatch between the number of paragraphs with NEs in the original GATE-annotated English texts (218) and in the translations produced by the three MT systems (225, 225 and 239 paragraphs with NEs). This is because the results of NE pre-processing could be submitted to the proprietary MT systems only in the form of a DNT list, which has its limitations. The most serious potential problem is over-generation: ambiguous items, which could be either NEs or common words in different contexts, are treated as NEs in *every* context, once they are written to the DNT list. For example, the word *Labour* could be either an organisation name ('the party'), a part of a larger NE, often of a type other than organisation name *(Federal Railway Labour Act),* or a common noun ('work', as in the phrase: *rise in labour costs).* As a result, in the translated corpus there are more NEs than in the original English corpus, annotated with GATE. This is reflected in the figures presented in row 1 of Table 2. Nevertheless, the difference is relatively low (less then 10% for the worst case). Given that there are (on average) only about 2 NE occurrences per paragraph in the corpus, over-generation does not greatly affect our evaluation results.

The segmentation method described above provided us with a clear formal distinction

between NE-internal and NE-external problems for MT. However, we made one exception to this distinction: in the DNT-processed English-French, Systran often incorrectly inserts definite articles for organisation names which are present in DNT list, but does not do so in the baseline translation. Our segmentation method treats these articles as part of the morphosyntactic context of NEs, and considerably increases the contextual degradation figures for Systran. But, linguistically, it is more correct to treat French articles as inner parts of NEs. Therefore, for the evaluation of contextual changes for Systran, we ignored strings where the inserted article was the only difference. As a result, Systran showed a net contextual improvement.

4. Results of the experiment

Table 5 summarises the results of the manual annotation of 50 strings containing differences for each MT system. (There are 61 scored differences for Systran, because in some strings there was more then one morphosyntactic or lexical difference).

Mark	ProMT 1998 E-R		ProMT 2001 E-F		Systran 2000 E-F	
	N	Score	N	Score	N	Score
+1*	28 =	+28.0	23 =	+ 23.0	18 =	+ 18.0
+0.5*	2 =	+1.0	5 =	+ 2.5	24 =	+ 12.0
0*	4 =	0	7 =	0	8 =	0
−0.5*	3 =	−1.5	1 =	− 0.5	1 =	− 0.5
−1*	13 =	−13.0	14 =	− 14.0	10 =	− 10.0
∑	50	+14.5	50	+ 11.0	61	+ 19.5
Gain		+29%		+22%		+32%

Table 5: Manual annotation results

N is the number of differences, annotated with that particular score. To compute the overall score for the system we multiplied the scores by the number of strings with this particular score, and added the results. The improvement was then computed by dividing the overall score by the number of scored differences: ∑score / ∑ N.

In order to see how the resulting scores change when more data is analysed, we continued scoring the English Russian ProMT 98 system, until 100 paragraphs with differences had been annotated. The results are presented in Table 6.

Mark	ProMT 1998 E-R	
	N	Score
+1*	59 =	+59.0
+0.5*	8 =	+4.0
0*	6 =	0
−0.5*	7 =	−3.5
−1*	31 =	−31.0
∑	111	+28.5
Gain		+26%

Table 6: Results for additional E-R data

We give an example of a sentence where improvement has been achieved in the DNT-processed translation for all three MT systems on several levels: morphological, syntactic and lexical.

	Original: The agreement was reached by a coalition of four *of Pan Am's* five unions.
E-R ProMT	**Baseline translation:** Соглашение было достигнуто коалицией четырех Кастрюли пять союзов Ама. ('The agreement was reached by a coalition of four of a Saucepan five unions of Am.') **DNT-processed translation:** Соглашение было достигнуто коалицией четырех из пяти союзов Pan Am. ('The agreement was reached by a coalition of four out of five unions *of Pan Am* ')
E-F ProMT	**Baseline translation:** L'accord a été atteint par une coalition de quatre de casserole cinq unions d'Am. ('The agreement was reached by a coalition of four of saucepan five unions of Am.') **DNT-processed translation:** L'accord a été atteint par une coalition de quatre de cinq unions de Pan Am. ('The agreement was reached by a coalition of four of five unions of Pan Am.')
E-F Systran	**Baseline translation:** L'accord a été conclu par une coalition de quatre de la casserole étais cinq syndicats. ('The agreement was reached by a coalition of four of the saucepan was five trades-unions.') DNT-processed translation: L'accord a été conclu par une coalition de quatre de Pan Am's cinq syndicats. ('The agreement was reached by a coalition of four of Pan Am's five trades-unions.')

Here are further typical cases of morphosyntactic improvement in the translated material:

Improved syntactic segmentation:

Original:
Representatives for the 5,400-member Allied *Pilots Association* didn't return phone calls.

E-R ProMT **Baseline translation:**
Представители для *Союзнических Пилотов с 5,400 членами Ассоциация* не возвращали обращения по телефону.
('Representatives for the *Allied Pilots with 5,400 members Association* didn't return phone calls.')

DNT-processed translation:
Представители для Allied Pilots Association с 5,400 членами не возвращали обращения по телефону.
Representatives for the *Allied Pilots Association* with 5,400-members didn't return phone calls.

Improved proper / common disambiguation:

Original:
A spokesman for the company said *American officials 'felt that...'*

E-F ProMT **Baseline translation:**
Un porte-parole de la société a dit que les fonctionnaires *américains* 'ont estimé que ...'
('A spokesman for the company said that the American [US] officials 'felt that... ")

DNT-processed translation:
Un porte-parole de la société a dit que les fonctionnaires *d'Américan* 'ont estimé que ...'
('A spokesman for the company said that the officials of American 'felt that...")

Improved word order:

Original:
USAir disclosed details of its plans for financing

E-F ProMT **Baseline translation:**
USAir les détails révélés de ses plans pour financer ...
('USAir the details revealed *(Pastparticiple)* of its plans for financing ... ')

DNT-processed translation:
USAir a révélé les détails de ses plans pour financer ...
('USAir revealed *(Verb)* the details of its plans for financing ...')

Improved lexical or syntactic disambiguation:

Original:
TWA *stock closed* at $28 ...

E-F Systran **Baseline translation:**
Fermé courant de TWA à $28 ...
('Closed *(Past participle)* current *(Noun/Present participle)* of TWA at $28 ... ')

DNT-processed translation:
L'action de TWA *s'est fermée à* $28 ...
('The stock of TWA closed *(Verb)* at $28 ...')

Original:
National Mediation Board is expected to release Pan Am Corp. and its Teamsters union from their long-stalled contract negotiations.

E-R ProMT **Baseline translation:**
Национальное Правление Посредничества, как ожидается, *выпустит* Кастрюлю - Корпорация и ее союз Водителей от их долго-остановленных переговоров контракта.
('National Mediation Board is expected to release [put on the market] a Saucepan - Corporation and its union of drivers from their long-stalled contract negotiations.')

DNT-processed translation:
National Mediation Board, как ожидается, *освободит* Pan Am Corp. И его союз Teamsters от их долго-остановленных переговоров контракта.
'National Mediation Board is expected to release [make free] Pan Am Corp. and its Teamsters union from their long-stalled contract negotiations.')

5. Conclusions

The results indicate that combining IE technology with MT has a great potential for improving the state-of-the art in output quality. Taking advantage of efforts to resolve specific linguistic problems - as has happened with NE recognition within the IE framework - improves not only the treatment of that phenomenon by MT, but also morphosyntactic and lexical well-formedness more generally in the wider context of the target, thus boosting the overall quality of MT. Our results show that modern MT systems still leave room to achieve a considerable improvement. Further gains in performance may be anticipated by harnessing other focussed technologies, such as word sense disambiguation, to MT.

We noted also that the scale of the improvement for particular MT systems correlates with the baseline quality of MT: it is more difficult to achieve improvement for a system which produces high-quality well-formed structures without DNT-processing. The improvement which is possible with NE DNT-processing is lowest for the English-French ProMT (Reverso) system. This system was ranked higher than English-French Systran by human evaluators in an experiment conducted by (Rajman and Hartley, 2001) using data from DARPA's 1992-1994 series of MT evaluations (White, et al, 1994). These human evaluations

confirmed the ranking predictions of an automatic evaluation algorithm which correlated the fluency, adequacy and informativeness scores awarded by human evaluators to the DARPA corpus with syntactic and semantic attributes of the corpus. In this respect, the measures of contextual improvement after DNT-processing with lists of NEs (organisation names) produced by GATE could be seen as a possible evaluation score for MT systems, which could lead to establishing a reliable quality scale for MT systems.

Future work will look at the sensitivity of the performance gain to corpus size and variation. Table 6 shows that the difference in the score for 50 annotated paragraphs and the score for 100 paragraphs for E-R ProMT98 is 3%. In general, different occurrences of the same NE tend to have a similar morphosyntactic context, so they constantly tend to either improve or worsen the quality. In a particular text, the same NEs tend to re-occur. As a result, an improvement or a decline in quality is usually not homogeneous across corpora, but is more constant for a particular text. The score changes in more or less homogeneous chunks of text. For E-R ProMT 98 MT system the average size of such chunks is about 7 differences (See Table 3, row 6 'Different strings per text'). For E-R ProMT 98, the value of each '+1' or '-1' score after 50 annotated differences is ±2%, so one text can potentially change the score by about ±14%. After checking 100 differences, the value of each '+1' or '-1' score becomes ±1%, so a new text could change the score by ±7% on average. In the case of E-R ProMT 98, scoring 50 additional new strings (about 7 new texts) changed the overall score by -3%. This indicates that, for our corpus, there is a reliable improvement after NE DNT-processing, but more work remains to be done.

Other future work will consider the well-formedness or acceptability of the NEs themselves.

References

Al-Onaizan, Y. and K. Knight. 2002. Translating Named Entities Using Monolingual and Bilingual Resources. *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics.* Philadelphia, July 2002. pp. 400-408.

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.

Cunningham, H., Y. Wilks and R. Gaizauskas. 1996. GATE -- a General Architecture for Text Engineering. *Proceedings of the 16th Conference on Computational Linguistics (COLING-96),* Copenhagen, Aug, 1996.

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.

Rajman, M. and T. Hartley. 2001. Automatically predicting MT systems ranking compatible with Fluency, Adequacy and Informativeness scores. *Proceedings of the 4th ISLE Workshop on MT Evaluation, MT Summit VIII.* Santiago de Compostela, September 2001. pp. 29-34.

White, J., T. O'Connell and F. O'Mara. 1994. The ARPA MT evaluation methodologies: evolution, lessons and future approaches. *Proceedings of the 1st Conference of the Association for Machine Translation in the Americas.* Columbia, MD, October 1994. pp. 193-205.