

# Translation Shared Task on Statistical Machine Translation: A Comparison of the Systems Output

Muhammad Arifur Rahman  
Assistant Professor  
Department of Physics  
Jahangirnagar University  
Dhaka, Bangladesh

## ABSTRACT

The ACL Workshop on Statistical Machine Translation proposed a translation shared task focused on European language pairs. Participants used their systems to translate into the target language a test set of unseen sentences in the source language. Involved languages were French, English, Spanish, German, Czech and Hungarian. The goal of this work is to quantitatively compare the translations generated by different systems. In particular, a selection of submitted runs for the French-English, German-English and Spanish-English tasks were analyzed. The systems involved in our investigation cover all the main approaches to machine translation, that is rule-based, statistical, example-based and hybrid.

## General Terms

Machine Learning, Machine Translation.

## Keywords

Shared Task, SMT, RBMT, EBMT, BLEU.

## 1. INTRODUCTION

Machine translation (MT) is the use of a mechanical device such as computer to translate texts or utterances of a natural language into another natural language. An MT system expects texts in a specific language as input and produces a text with a corresponding meaning in a different language as output. MT can be seen as a decision problem where we have to decide on the best of target language text matching a source language text.

Automatic MT evaluation is a mean of scoring the output from an MT system with respect to a set of human reference translations. The basic principle is that a translation is good if it is close to a member of such a set. People are interested in this measure because it can provide immediate feedback compared with human judgments, although its correlation with human evaluation of accuracy and fluency is not perfect and still under investigation.

Recent work on statistical MT (SMT) has lead to a significant progress in coverage and quality of translation technology. At present time MT systems mainly follow three different approaches: statistical, rule based and example based. Phrase-based SMT systems are trained from the parallel corpora [1] and originated from the IBM models [2]. Rule based machine translation (RBMT) systems rely on rules and linguistic resources built for that purpose. Example based machine translation (EBMT) is comparatively a new approach, used by decoding knowledge from a bilingual text. Chen et al., (2007) describes an architecture that allows to combine SMT with

one or multiple RBMT systems in a multiengine setup which can be termed as a type of hybrid system[3].

The shared task of ACL-2008 Workshop on SMT proposed shared task translation between 10 pairs of European languages. Around thirty systems from different universities and companies participated to the task for different language pairs [4]. This report presents an analysis of the output of various systems participated to the French to English, Spanish to English and German to English language pair tasks. The system comparison is made as defined in the following section.

We ask that authors follow some simple guidelines. In essence, we ask you to make your paper look exactly like this document. The easiest way to do this is simply to download the template, and replace the content with your own material.

## 2. COMPARISON METHOD

This work aims at comparing the translations generated by the systems that participated to the shared task. The rationale behind this investigation is to verify if systems sharing the approach, rule-based, example-based or statistical, also provide similar outputs. The comparison is quantitative and made by computing the score of an automatic evaluation metric, the BLEU score (see Section 2.1), for each pair of systems, taking as hypothesis the output from one system and as reference the output of the other system. Once the scores are available for each pair, the matrix collecting them can be analyzed for finding regularities in the behavior of MT systems.

### 2.1 The BLEU Score

BLEU was developed by Kishore Papineni and others in 2002 [5]. Papineni's argument is: how does one measure translation performance? One of the answers they provided is: the closer the MT is to a professional human translation, the better it is. BLEU is an automatic evaluation metric based on the matching of n-gram of words. To compute the BLEU score, one has to count the number of n-grams in the hypothesized translation that have a match in the corresponding reference translation. That is the precision  $p_n$  is computed as:

$$p_n = \frac{\sum_{h \in TS} \sum_{n\text{-gram} \in h} \text{Count}_{\text{matched}}(n\text{-gram})}{\sum_{h \in TS} \sum_{n\text{-gram} \in h} \text{Count}(n\text{-gram})}$$

where h is the hypothesis,  $\text{Count}_{\text{matched}}(n\text{-gram})$  is the count of n-grams found both in the hypothesis and reference.

Count(n-gram) is the count of n-grams in the hypothesis and TS is the test set. Typically, precisions are computed for n=1..4.

Multiple reference translations are also used to increase accuracy where paraphrases may exist. In this case, BLEU uses a modified n-gram precision. This precision clips the count for each n-gram in any test translation to prevent it from exceeding the count of this n-gram in the best matching reference translation.

Since BLEU is precision based and recall is difficult to formulate over multiple reference translations, a brevity penalty is introduced to compensate for the possibility of proposing high precision hypothesis translations which are too short. IBM's formula for calculating the BLEU score is [6]:

$$BLEU_{SCORE} = BP_{BLEU} \times PRECISION_{BLEU}$$

where the precision factor is the geometric mean of single n-gram precision:

$$PRECISION_{BLEU} = \exp\left(\sum_n \frac{1}{N} \log(p_n)\right)$$

while the brevity penalty is calculated as:

$$BP_{BLEU} = \begin{cases} 1 & \text{if } LenH > LenR \\ \exp\left(1 - \frac{LenR}{LenH}\right) & \text{if } LenH \leq LenR \end{cases}$$

where LenH is the length of the corpus of hypothesis translations, and LenR is the effective reference corpus length. In case of multiple references, it is calculated as the sum of the single reference translation from each set which is closest to the hypothesis translation.

One of the main limitations of the BLEU is that since the evaluation is simply based on n-gram matching, it is possible to give the same score to sentences with completely different meaning by switching words/n-grams. Moreover in the calculation of the BLEU score, longer n-grams dominate the shorter n-grams.

### 3. DESCRIPTIONS OF SYSTEMS

In this section, a brief description of each of the systems participating to the shared task is provided; first, however, an overview of the different approaches to MT is supplied.

#### 3.1 Rule Based Machine Translation

RBMT needs linguistic knowledge to create appropriate rules of translation. RBMT paradigm is associated with systems that rely on different linguistic levels of rules for translation between the source and the target language. The prototypical example is Rosetta [7], an interlingual system which divides translation rules in two categories [8]. One is "M-rules", which are "meaning-preserving rules", that is map's between syntactic trees and underlying meaning structures, the other is "S-rules", which are "non-meaningful rules", that is map's lexical items and syntactic trees.

In RBMT, a grammar is comprised with a lot of rewriting rules [9-10]. Translation is carried out by repeating pattern matching and transformation of tree or graph structures that represent the syntax or semantics of a sentence. A great part of the processing time is spent in pattern matching which mostly results in failure. The key to improve the processing efficiency is how to avoid the pattern matching that results in failure. A number of methods such as the Rete pattern match algorithm [11] have been developed.

#### 3.2 Example Based Machine Translation

EBMT is about how to decode knowledge from bilingual texts, where the knowledge seems to have no overt formal representation or any encoding scheme. The EBMT approach became popular soon after some positive results were published in a number of papers demonstrating its plausibility. Sato and Nagao [12] investigated the problem of example selection by approximate (or inexact) matching of input sentences and example sentences, using a similarity measure. By around 1993, EBMT had become an established research field of MT and many example-based techniques were applied to various MT tasks. [13] attempted the example-based translation of computer technical terms with respect to the focus term and its surrounding contexts.

An example is a pair (or couple) of texts in two languages that are a translation of each other. The texts can be of any size at any linguistic unit: words, phrase, clause, sentence, and paragraph. A critical issue that needs to be examined closely in this context is the number of examples over a large-scale bilingual corpus, which can be unlimited in practice. The example number can be exponentially large in respect to the corpus size, if all possible examples from a bilingual corpus are selected. Consequently, the problems of scale of EBMT might arise, because any fragment of a sentence can be an example.

#### 3.3 Statistical Machine Translation

Pure SMT is relatively a new research area. SMT has been introduced by the research group at IBM [14] in 1990. They introduced the concept of alignment models to describe the dependencies between source and target language words [2] and developed a search algorithm for these models based on the paradigm of stack decoding. A SMT model is a mathematical model in which the process of human language translation is statistically modeled [15]. Model parameters are automatically estimated using a corpus of translation pairs. Unfortunately, even for simple translation models, the search problem is NP complete. Various research groups tried to extend the IBM work to develop more efficient search algorithms by using suitable simplifications and applying better optimization methods. A major disadvantage of the baseline IBM alignment models is that they do not take word context into account. A partial solution to this problem is phrase based SMT [1], where word sequences (phrases) are translated rather than single words, providing SMT with a robustness in word selection and local word reordering.

#### 3.4 Hybrid Systems

The traditional RBMT system is expensive in terms of formulating rules. It easily introduces inconsistencies, and it is too rigid to be robust. In contrast, the SMT system is based on noisy channel model and is robust in processing partial and ill-formed sentences. The computation time in processing long

**Table 1: Participants in the shared translation task with basic approaches and official BLEU score on Test2008 data**

ID	Participant	Basic Approach	Language Pair	Official BLEU score
CMU-SMT	Carnegie Mellon University SMT	SMT based on Weighted Sentence Pair [19]	Es-En	33.23
			De-En	17.00
CMU-STATXFER	Carnegie Mellon University StatXFER	Syntax Based Translation System [20]	Fr-En	20.00
CUED	Cambridge University	SMT- follows the Transducer Translation model [16]	Fr-En	32.83
			Es-En	33.11
DCU	Dublin City University	Hybrid System: Combination of EBMT and SMT [22]	Fr-En	31.63
			Es-En	32.74
LIMSI	LIMSI	SMT based on Moses Toolkit [17]	Fr-En	33.00
			ES-En	33.00
			De-En	27.00
LIU	Link'oping University	SMT based on POS factor and Sequential model on POS [26]	De-En	27.00
LIUM-SYSTRAN	University of Le Mans and Company Systran	Hybrid system: Combination of SMT based on Moses Toolkit and RBMT [23]	Fr-En	33.13
RBMT1-6	Babelfish, Lingenio, Lucy, OpenLogos, ProMT, SDL (ordering anonymized)	RBMT	Fr-En	17.00 to 20.00
			Es-En	17.00 to 19.00
			De-En	12.00 to 15.00
SAAR	University of Saarbruecken	Hybrid System: Multi Engine with Moses SMT [24]	Fr-En	28.12
			Es-En	33.15
			De-En	27.00
SYSTRAN	Systran	RBMT [21]	Fr-En	26.00
UCL	University College London	Hybrid System: Phrased-based SMT with Kernel Regression Model [25]	Fr-En	26.59
UEDIN	University of Edinburgh	SMT based on Moses Toolkit [18]	Fr-En	33.50
			Es-En	33.70
			De-En	28.00
UPC	Universitat Politecnica de Catalunya, Barcelona	N-gram Based SMT [27]	Es-En	32.88

sentences sharply increases as the number of words increases. The EBMT heavily depends on the quality of collected examples and the similarity measures between examples and input sentences. When the matched units are subsentential structures (e.g., phrase structures), the performance of such a system is better than that of a word-level system. As for the knowledge-based system the difficulties are in how the knowledge is represented, how fine the knowledge is, and what the inference engine is. In addition, the cost of compiling knowledge is expensive.

The hybrid design is an effort to produce a method that makes the most of the strengths of above mentioned methods and that compensate for their weaknesses. At AMTA 2004 and MT Summit 2005 just about all commercial MT developers also claimed to have hybrid systems. Much current research in MT is neither based purely on linguistic knowledge nor on statistics, but includes some degree of hybridization.

The hybridization of MT approaches attempted so far, primarily concentrates on two aspects: the technical aspect and the improved translation quality. A prominent example is

a parallel run of two or more different MT engines and the combination of their output.

### 3.5 Systems participating in Shared Task

Table 1 lists the participants, to the translation shared task whose runs we have investigated. A short description of the systems are given below. The **Cambridge University** Engineering Department SMT system follow the Transducer Translation Model, a phrase based generative model of translation that applies a series of transformation specified by a conditional probability distribution and encoded as Weighted Finite State Transducers [16]. **LIMSI** used an SMT system based on Moses toolkit. For Europarl, n-best rescoring was performed using an enhanced n-gram or a neuronal language model [17]. **University of Edinburgh**'s system used the Moses decoder which follow the phrase based SMT approach, with default setting as a starting point. They added minimum Bayes risk decoding and reordering constraints to the decoder [18]. For their SMT system, **Carnegie Mellon University SMT** applied the methods of weighting sentence pair using language models, and extended the general weighting method to genre-dependent weight [19]. The **Carnegie Mellon University Statistical Transfer** (Stat-XFER) is a transfer engine using two language-pair-dependent resources: a grammar of weighted synchronous context-free rules, and a probabilistic bilingual lexicon of syntax-based word and phrase level translation [20]. **RBMT1** to **RBMT6** are Rule Based Machine translation systems from six companies, namely Babelfish, Lingenio, Lucy, OpenLogos, ProMT, SDL with anonymized ordering [4]. The system of **Systran** was a rule based system that in recent years began integrating statistical features and corpus based model [21]. **Dublin City University's MATREX** (Machine Translation using Examples) is a data-driven MT. It is a hybrid system which exploits EBMT and SMT techniques to build a combined translation model [22]. LIUM/Systran is a joint project between the **University of Le Mans** and company **Systran**. It is a hybrid system. They combine the data driven approaches and the concentrated knowledge present in RBMT resources [23]. **University of Saarbruecken** uses a hybrid system of SMT and RBMT in a

multi engine setup. It is a variant of standard SMT technology to align translation from one or more RBMT system with the source text and incorporated phrases extracted from these alignments into the phrase table of the SMT system [24]. It uses Moses to find good combination of phrases from SMT training data with the phrases derived from RBMT. **University College London**'s system uses a hybrid system of classic phrase-based SMT model and the kernel regression model [25]. First, for each source sentence a small relevant set of sentence pairs are retrieved from the large scale parallel corpus. Then the regression model is trained on this small relevant set only as a sparse approximation of the regression hyperplane trained on the entire training set.

## 4. RESULTS

The first part of this section provides a brief description of data used for the experiments while the second part provides the findings of experiments and a brief description of results for different systems.

### 4.1 Data

For our investigation we have compared the automatic translations provided by the systems on the Europarl test2008 data set. Table 2 shows the test set statistics.

Table 2: Statistics of Europarl Test 2008 data

# Sentences	2,000
# Words	60,185
# Distinct Words	6,050

### 4.2 Experiments

Table 3 shows intra-system Bleu scores computed on the outputs of the French-to-English task. It can be noted that higher scores refer to pairs of pure statistical-based systems (CUED, LIMSI and UEDIN) and hybrid systems which include a statistical component (LIUM-SYSTRAN, DCU). It is not surprising to observe that the highest score corresponds to the pair of Moses-based systems (UEDIN and LIMSI). Anyway, also CUED, which is an SMT system not based on Moses, shows high scores with the other SMT systems: this should mean that the similarity of translations from statistical systems is due to the approach rather than the actual implementation.

Table 3: French to English Europarl Test2008 Data

Uedin	0.412	0.6955	0.7028	0.7867	0.7689	0.3328	0.3363	0.3182	0.5255	0.531	0.4491	1
Ucl	0.2629	0.4496	0.4282	0.4403	0.4428	0.1879	0.1993	0.1786	0.4116	0.3408	1	0.4496
Systran	0.3724	0.508	0.5105	0.5411	0.5371	0.3734	0.3776	0.37	0.4721	1	0.3404	0.531
Saar	0.3046	0.498	0.4891	0.524	0.5205	0.2791	0.2948	0.2679	1	0.4699	0.4122	0.523
Rbmt6	0.3272	0.2996	0.3192	0.3319	0.3287	0.4835	0.3899	1	0.2724	0.3696	0.1793	0.3178
Rbmt4	0.2934	0.3183	0.3323	0.3462	0.3437	0.4158	1	0.3903	0.298	0.3771	0.1995	0.3359
Rbmt3	0.3506	0.3183	0.3314	0.349	0.346	1	0.4153	0.4831	0.285	0.3732	0.189	0.3326
lium-systran	0.4256	0.6808	0.6971	0.7919	1	0.346	0.3442	0.329	0.5225	0.5373	0.4422	0.7691
Limsi	0.432	0.6897	0.7183	1	0.792	0.3491	0.3467	0.3322	0.5259	0.5414	0.4396	0.787
Dcu	0.4136	0.6429	1	0.7185	0.6973	0.3314	0.3327	0.3194	0.4907	0.5108	0.4275	0.7032
Cued	0.3966	1	0.6426	0.6895	0.6807	0.3184	0.3187	0.2999	0.5002	0.5081	0.449	0.6956
cmu-statxfer	1	0.3961	0.4132	0.4315	0.4251	0.351	0.2931	0.3273	0.3093	0.372	0.2637	0.4115
	cmu-statxfer	Cued	dcu	limsi	lium-systran	rbmt3	rbmt4	rbmt6	Saar	systran	ucl	uedin

**Table 4: Spanish to English Europarl Test2008 Data**

<b>Upc</b>	0.6511	0.6353	0.6074	0.6432	0.3551	0.3471	0.3373	0.5979	0.4069	0.6302	1
<b>Uedin</b>	0.8008	0.6781	0.7168	0.7688	0.3273	0.3243	0.3129	0.7046	0.454	1	0.6304
<b>Ucl</b>	0.4455	0.4371	0.4274	0.4375	0.1954	0.1899	0.1831	0.4327	1	0.4541	0.4069
<b>Saar</b>	0.699	0.6397	0.64	0.669	0.3798	0.3765	0.3809	1	0.4322	0.7052	0.5985
<b>Rbmt6</b>	0.3306	0.3252	0.327	0.3413	0.522	0.5303	1	0.3808	0.1845	0.3127	0.3369
<b>Rbmt4</b>	0.3383	0.3347	0.3338	0.3508	0.5308	1	0.5301	0.3766	0.1917	0.3241	0.3468
<b>Rbmt3</b>	0.3428	0.3408	0.3403	0.3542	1	0.5311	0.5221	0.3797	0.1968	0.327	0.3547
<b>Limsi</b>	0.7512	0.6751	0.7429	1	0.3545	0.3509	0.3415	0.6686	0.4372	0.7691	0.6436
<b>Dcu</b>	0.7036	0.6449	1	0.7427	0.3406	0.334	0.3272	0.6395	0.4272	0.7168	0.6077
<b>Cued</b>	0.684	1	0.6451	0.6751	0.3411	0.3348	0.3254	0.6393	0.4368	0.6783	0.6357
<b>cmu-smt</b>	1	0.6841	0.7039	0.7514	0.343	0.3383	0.3307	0.6987	0.4451	0.8012	0.6516

**cmu-smt**    **Cued**    **dcu**    **limsi**    **rbmt3**    **rbmt4**    **rbmt6**    **Saar**    **ucl**    **uedin**    **upc**

**Table 5: German to English Europarl Test2008 Data**

<b>Uedin</b>	0.3319	0.5819	0.5755	0.2117	0.2655	0.265	0.2583	0.2404	0.5718	0.317	1
<b>Ucl</b>	0.2107	0.3399	0.3403	0.1314	0.1387	0.1336	0.1317	0.1232	0.3354	1	0.3153
<b>Saar</b>	0.3157	0.598	0.65	0.2169	0.2896	0.2649	0.2571	0.2626	1	0.337	0.5719
<b>rbmt6</b>	0.1944	0.2196	0.2272	0.3	0.3651	0.3733	0.3968	1	0.2622	0.1245	0.2401
<b>rbmt5</b>	0.2025	0.2357	0.2412	0.3028	0.3855	0.4109	1	0.3969	0.2568	0.133	0.258
<b>rbmt4</b>	0.1973	0.2383	0.2459	0.31	0.407	1	0.4109	0.3733	0.2646	0.135	0.2647
<b>Rbmt3</b>	0.2066	0.2424	0.2468	0.3117	1	0.4072	0.3857	0.3653	0.2893	0.1399	0.2652
<b>Rbmt1</b>	0.1518	0.1944	0.2005	1	0.309	0.3069	0.2998	0.2969	0.2156	0.1306	0.2104
<b>Liu</b>	0.3231	0.6172	1	0.2015	0.2471	0.2462	0.2415	0.2275	0.6503	0.3415	0.5759
<b>Limsi</b>	0.3289	1	0.6173	0.1954	0.2427	0.2386	0.236	0.2199	0.5983	0.341	0.5823
<b>Cmu-statxfer</b>	1	0.3285	0.3227	0.153	0.2066	0.1975	0.2026	0.1946	0.3155	0.2123	0.3317

**cmu-statxfer**    **Limsi**    **Liu**    **rbmt1**    **rbmt3**    **rbmt4**    **rbmt5**    **rbmt6**    **saar**    **ucl**    **uedin**

On the other side, the scores of pairs where at least one system is rule or example-based vary greatly and are quite lower than scores of pairs of systems based on statistics. For example, the BLEU scores computed on any translation of the three commercial RBMT systems do not exceed 48.31, the maximum corresponding to the pair RBMT3 and 6, and typically is below 40. Also SYSTRAN output is quite different from any other translation, its BLEU scores ranging in the 34-54 interval. Interestingly, the LIUM-SYSTRAN system seems to generate translations similar to the other statistical-based systems (68-79) but not to the SYSTRAN output (53.7).

Tables 4 and 5 are the analogous of Table 3 for the Spanish-English and German-English tasks, respectively. On the basis of reported figures, quite similar considerations can also be made for these tasks. Note that German-English scores are lower than those of the other two tables; this could be due to the fact that since this is a more difficult task, systems are wrong more often and errors, by their own nature, are different because randomly investigated.

## 5. CONCLUSION

In this work we have compared some of the runs submitted to the translation shared task of the ACL 2008 Workshop on

SMT. The investigation regarded three language pairs, French-English, German-English and Spanish-English, and systems implemented following different approaches to MT, namely the rule-based, the statistical, the example-based and the hybrid one. The quantitative comparison was based on the BLEU score.

A clear outcome of our analysis is that statistical-based systems generate similar translations, whatever their implementation; in fact, high BLEU scores (denoting high correlation) were observed for pairs involving not only Moses-based systems, but also other statistical systems. Similar translations are generated also by hybrid systems which include a statistical component.

On the contrary, rule- and example-based systems do not show such a behavior. The BLEU scores measured for translation pairs where at least one output is from a non-statistical system are quite low. It is worth to notice that this is true also for Systran and commercial systems, showing that on the market one can expect to find MT systems that differ significantly.

An interesting result of our work regards system combination. It is known that when translations from different systems are combined with the goal of improving the overall quality, it is convenient that translations differ as much as possible. Our study suggests not forget to include outputs from non-statistical engines.

## 6. REFERENCES

- [1] Philipp Koehn, Franz Josef Och, and Daniel Marcu. 2003. Statistical Phrase-based translation. In Proceedings of Human Language Technology Conference of the North American Chapter of the Association for Computational Linguistics (HLT-NACCL, pages 127-133. Edmonton, Alberta.
- [2] 2 Brown, P. F., Pietra, S. D., Pietra, V. J. D., and Mercer, R. L., 1993. The mathematic of statistical machine translation: Parameter estimation. *Computational Linguistics*, 19(2):263–311.
- [3] Yu Chen, Andreas Eisele, Christian Federmann, Eva Hasler, Michael Jellinghaus and Silke Theison. 2007. Multiengine Machine translation with an open-source SMT decoder. In *Proceedings of WMT 07*, pages 193-196, Prague, Czech Republic, June. Association for Computational Linguistics.
- [4] Chris Callison-Burch, Cameron Fordyce, Philipp Kohen, Christof Monz and Josh Schroeder. 2008. Further Meta-Evaluation of Machine translation. In *Proceedings of WMT 08*, pages 70-106, Columbus, Ohio, USA, June. Association for Computational Linguistics.
- [5] Kishore Papineni, Salim Roukos, Todd Ward and Wei-Jing Zhu. 2002. BLEU: a method for automatic evaluation of machine translation. In *proceedings of the 40<sup>th</sup> Annual Meeting of the Association for Computational Linguistics (ACL-2002)*. Philadelphia, Pennsylvania.
- [6] Andrew Finch, Young-Sook Hwang and Eiichiro Sumita, "Using Machine Translation Evaluation Techniques to Determine Sentence-level Semantic Equivalence", pages 17-24 Second International Joint Conference, Jeju Island, Korea.
- [7] Rosetta, M. T. 1994. Compositional translation. International Series in Engineering and Computer Science, Kluwer Academic Publishers.
- [8] Dorr Bonnie J., Pamela W. Jordan, and John W. Benoit, 1998. A Survey of Current Paradigms in Machine Translation," Technical Report: LAMP-TR-027, UMIACS-TR-98-72, CS-TR-3961, University of Maryland, College Park.
- [9] Boitet, C. Ch. Boitet, P. Guillaume, M. Quezel-Ambrunaz, 1982. Implementation and conversational environmerz of ARIANE-78.4. In *Proceedings of COLING-82*, Prague.
- [10] Jun-tcht Nakamura, Jun-tcht Tsujii, Makoto Nagao, 1984. Grammar Writing System (GRADE) of Mu-Machtne Translation Project and its Characteristics. In *Proceedings of COLING-84*, pages: 338-343, Stanford University, Stanford, California.
- [11] Forgy, C. L. 1982. Rete : A fast algoritm for the many pattern/many object pattern match problems. September 1982, pages: 17-37. Artificial Intelligence, Vol 19(1).
- [12] Sato, S. and Nagao, M. 1990. Towards memory based translation. In *Coling-90: papers presented to the 13th International Conference on Computational Linguistics*, volume 3, pages 173–180.
- [13] Sato, S. 1993. Example-based translation of technical terms. In TMI-93: The Fifth International Conference on Theoretical and Methodological Issues in Machine Translation.
- [14] Peter F. Brown, John Cocke, Stephan A. Della Pietra, Vincent J. Della Pietra, Fredrick Jelinek, John D. Lafferty, Robert L. Mercer and Paul S. Roossin. 1990. A statistical approach to machine translation. *Computational Linguistics*, 16(2):79-85.
- [15] Yamada, K. and Knight, K., 2001. A syntax-based statistical translation mode. In *Proceedings of Association of Computational Linguistics*.
- [16] Graeme Blackwood, Adri`a de Gispert, Jamie Brunning, andWilliam Byrne. 2008. European language translation with weighted finite state transducers: The CUED MT system for the 2008 ACL workshop on SMT. In *Proceedings of the Third Workshop on Statistical Machine Translation*, pages 131–134, Ohio, June. Association for Computational Linguistics.
- [17] Daniel D´echelotte, Gilles Adda, Alexandre Allauzen, H´el`ene Bonneau-Maynard, Olivier Galibert, Jean-Luc Gauvain, Philippe Langlais, and Francois Yvon. 2008. Limsi’s statistical translation systems for WMT’08. In *Proceedings of the Third Workshop on Statistical Machine Translation*, pages 107–110, Columbus, Ohio, June. Association for Computational Linguistics.
- [18] Philipp Koehn, Abhishek Arun, and Hieu Hoang. 2008. Towards better machine translation quality for the German-English language pairs. In *Proceedings of the Third Workshop on Statistical Machine Translation*, pages 139–142, Columbus, Ohio, June. Association for Computational Linguistics.
- [19] Nguyen Bach, Qin Gao, and Stephan Vogel. 2008. Improving word alignment with language model based confidence scores. In *Proceedings of the Third Workshop on Statistical Machine Translation*, pages 151–154, Columbus, Ohio, June. Association for Computational Linguistics.
- [20] Greg Hanneman, Edmund Huber, Abhaya Agarwal, Vamshi Ambati, Alok Parlikar, Erik Peterson, and Alon Lavie. 2008. Statistical transfer systems for French-English and German-English machine translation. In *Proceedings of the Third Workshop on Statistical Machine Translation*, pages 163–166, Columbus, Ohio, June. Association for Computational Linguistics.
- [21] Lo`ic Dugast, Jean Senellart, and Philipp Koehn. 2008. Can we relearn an RBMT system? In *Proceedings of the Third Workshop on Statistical Machine Translation*, pages 175–178, Columbus, Ohio, June. Association for Computational Linguistics.
- [22] John Tinsley, Yanjun Ma, Sylwia Ozdowska, and Andy Way. 2008. MaTrEx: The DCU MT system for WMT 2008. In *Proceedings of the Third Workshop on Statistical Machine Translation*, pages 171–174, Columbus, Ohio, June. Association for Computational Linguistics.
- [23] Holger Schwenk, Jean-Baptiste Fouet, and Jean Senellart. 2008. First steps towards a general purpose French/English statistical machine translation system. In *Proceedings of the Third Workshop on Statistical Machine Translation*, pages 119–122, Columbus, Ohio, June. Association for Computational Linguistics.

- [24] Andreas Eisele, Christian Federmann, Hervé Saint-Amand, Michael Jellinghaus, Teresa Herrmann, and Yu Chen. 2008. Using Moses to integrate multiple rule-based machine translation engines into a hybrid system. In *Proceedings of the Third Workshop on Statistical Machine Translation*, pages 179–182, Columbus, Ohio, June. Association for Computational Linguistics.
- [25] Zhuoran Wang and John Shawe-Taylor. 2008. Kernel regression framework for machine translation: UCL system description for WMT 2008 shared translation task. In *Proceedings of the Third Workshop on Statistical Machine Translation*, pages 155–158, Columbus, Ohio, June. Association for Computational Linguistics.
- [26] Sara Stymne, Maria Holmqvist, and Lars Ahrenberg. 2008. Effects of morphological analysis in translation between German and English. In *Proceedings of the Third Workshop on Statistical Machine Translation*, pages 135–138, Columbus, Ohio, June. Association for Computational Linguistics.
- [27] Maxim Khalilov, Adolfo Hernández H., Marta R. Costajussa, Josep M. Crego, Carlos A. Henríquez Q., Patrik Lambert, José A. R. Fonollosa, José B. Mariño, and Rafael E. Banchs. 2008. The TALP-UPC Ngram-based statistical machine translation system for ACLWMT 2008. In *Proceedings of the Third Workshop on Statistical Machine Translation*, pages 127–130, Columbus, Ohio, June. Association for Computational Linguistics.