

# Language-independent hybrid MT with PRESEMT

**George Tambouratzis**  
ILSP, Athena R.C  
[giorg\\_t@ilsp.gr](mailto:giorg_t@ilsp.gr)

**Sokratis Sofianopoulos**  
ILSP, Athena R.C  
[s\\_sofian@ilsp.gr](mailto:s_sofian@ilsp.gr)

**Marina Vassiliou**  
ILSP, Athena R.C  
[mvas@ilsp.gr](mailto:mvas@ilsp.gr)

## Abstract

The present article provides a comprehensive review of the work carried out on developing PRESEMT, a hybrid language-independent machine translation (MT) methodology. This methodology has been designed to facilitate rapid creation of MT systems for unconstrained language pairs, setting the lowest possible requirements on specialised resources and tools. Given the limited availability of resources for many languages, only a very small bilingual corpus is required, while language modelling is performed by sampling a large target language (TL) monolingual corpus. The article summarises implementation decisions, using the Greek-English language pair as a test case. Evaluation results are reported, for both objective and subjective metrics. Finally, main error sources are identified and directions are described to improve this hybrid MT methodology.

## 1 Introduction and background

Currently a large proportion of language-independent MT approaches are based on the statistical machine translation (SMT) paradigm (Koehn, 2010). A main benefit of SMT is that it is directly amenable to new language pairs, provided appropriate training data are available for extracting translation and language models. The main obstacle to the creation of an SMT system

is the requirement for SL-TL parallel corpora of a sufficient size to allow the extraction of meaningful translation models. Such corpora (of the order of million sentences) are hard to obtain, particularly for less resourced languages. On the other hand, the translation accuracy of such systems largely depends on the quality and size of the bilingual corpora, as well as their relevance to the domain of text being translated. Even if such parallel corpora exist for a language pair, they are frequently restricted to a specific domain (or a narrow range of domains). As a consequence, these corpora are not suitable for creating MT systems that focus on other domains. For this reason, in SMT, researchers are investigating the extraction of information from monolingual corpora, including lexical translation probabilities (Klementiev et al., 2012) and topic-specific information (Su et al., 2011).

Alternative techniques for creating MT systems using less informative but readily available resources have been proposed. Even if these methods do not provide a translation quality as high as SMT, their ability to develop hybrid MT systems with very limited specialised resources represents an important advantage. Such methods include automatic inference of templates for structural transfer from SL to TL (Caseli et al., 2008 and Sanchez-Martinez et al., 2009). Similarly, Carbonell et al. (2006) propose an MT method that needs no parallel text, but relies on a lightweight translation model utilising a full-form bilingual dictionary and a decoder for long-range context. Other systems using low-cost resources include METIS (Dologlou et al., 2003) and METIS-II (Markantonatou et al., 2009; Carl et al., 2008), which utilise a bilingual lexicon

and monolingual corpora to translate SL texts. METIS/METIS II, which have studied translation only towards English, employ pattern recognition algorithms to retrieve the most appropriate translation from a monolingual corpus.

## 2 The MT methodology in brief

The MT methodology has been developed within the PRESEMT (Pattern REcognition-based Statistically Enhanced MT) project, funded by the European Commission (cf. [www.presemt.eu](http://www.presemt.eu)). It comprises three stages:

- (i) pre-processing, where the input sentence is tagged and lemmatised
- (ii) main translation, where the actual translation output is generated and
- (iii) post-processing, where the corresponding tokens are generated from lemmas.

The main translation process is split in two phases, namely (a) the establishment of the translation structure in terms of phrase order and (b) the definition of word order and resolution of lexical ambiguities at an intra-phrase level.

In terms of resources, PRESEMT utilises a bilingual lemma dictionary providing SL – TL lexical correspondences. It also employs an extensive TL monolingual corpus, compiled automatically via web crawling (Pomikalek et al., 2008) to generate a comprehensive phrase-based language model. The provision of the monolingual corpus allows PRESEMT to use only a very small bilingual corpus for mapping the transfer from SL to TL sentence structures. This bilingual corpus only numbers a few hundred sentences, reducing reliance on costly linguistic resources. The corpus is assembled from available parallel corpora, only replacing free translations with more literal ones, to allow the accurate extraction of structural modifications. The parallel corpus coverage is not studied prior to integration in PRESEMT, which would have allowed an optimisation of translation performance.

## 3 Extracting information from corpora

### 3.1 Parallel corpus

Initially, both the bilingual and the monolingual corpora are annotated<sup>1</sup> so as to incorporate lemma and Part-of-Speech (PoS) information and other salient language-specific morphological features (e.g. case, number, tense etc.). Furthermore, for the TL side, a shallow parser or chunker (hereafter referred to as parser) is used to split the sentences into syntactic phrases. As the proposed methodology has been developed to maximise the use of publicly-available software, the user is free to select any desired parser for the TL language.

To avoid either an additional SL side parser or potential incompatibilities between the two parsers, the Phrase Aligner module (**PAM**, Tambouratzis et al., 2011) is implemented. PAM transfers the TL side parsing scheme, which encompasses lemma, tag and parsing information, to the SL side, based on lexical information coupled with statistical data on PoS tag correspondences extracted from the lexicon. The parsing scheme includes phrase boundaries and phrase labels. PAM follows a 3-step process, involving (a) lexicon-based alignment, (b) alignment based on similarity of grammatical features and PoS tag correspondence and (c) alignment on the evidence of already aligned neighbouring words.

The SL side of the aligned corpus is subsequently processed by the Phrasing model generator (**PMG**), to create an SL phrasing model which will then parse sentences input for translation. The original PMG implementation (Tambouratzis et al., 2011) has utilised Conditional Random Fields (CRF), due to the considerable representation capabilities of this model (Lafferty et al., 2001). CRF is a statistical modelling method that takes context into account to predict labels for sequences of input samples.

The implementation of an alternative PMG methodology (termed PMG-simple) based on template-matching principles has also been pursued. PMG-simple locates phrases that match

---

<sup>1</sup> For the annotation task readily available tools are employed. For the experiments reported here, TreeTagger (Schmid, 1994) has been used for the TL text processing and the FBT PoS tagger (Prokopidis et al., 2011) has been employed for the processing of the SL text..

exactly what it has seen before, based on a simple template-matching algorithm (Duda et al., 2001). The templates used are the phrases to which the SL side sentences of the bilingual corpus have been segmented. In contrast to CRF, PMG-simple implements a greedy search (Black, 2005) without backtracking. Initially all phrases are positioned in an ordered list according to their likelihood of being accurately detected. Starting from the phrase with the highest likelihood, PMG-simple examines if each phrase occurs in the input sentence. If it does and the constituent words are not part of an already established phrase, the constituent words are marked as parts of this phrase and are no longer considered in the phrase-matching process. If the phrase pattern does not occur, the next in-line phrase is considered, until the table is exhausted. Comparative results between CRF and PMG-simple are reported in the results section.

### 3.2 Monolingual corpus

The TL monolingual corpus is processed to extract two complementary types of information. The first type supports disambiguation between multiple possible translations, while the second determines the order of words in the final translation and the addition or removal of functional words, using a TL phrase model derived from an indexing based on (i) phrase type, (ii) phrase head lemma and (iii) phrase head PoS tag.

The TL phrases are then organised in a hash map that allows the storage of multiple values for each key, using as a key the three aforementioned criteria. For each phrase the number of occurrences within the corpus is retained. Each hash map is stored in a separate file to minimise access time during translation.

## 4 Translation phase 1: Structure selection

The Structure selection phase determines the type and relative position of TL phrases to which the SL ones are translated. To achieve this, PRESEMT consults the SL-to-TL structural modifications as contained in the PAM-processed parallel corpus. In that respect, it resembles EBMT (Hutchins, 2005).

Translation phase 1 receives as input an SL sentence, annotated with tag & lemma informa-

tion and segmented into phrases by the PMG. A dynamic programming algorithm then determines for each SL side the most similar (in terms of phrase structure) SL sentence from the bilingual corpus. Similarity is calculated by taking into account structural information such as the phrase type, the PoS tag and case (if applicable) of the phrase head and phrase functional head info. The phrases of the input sentence are then reordered to generate the translation structure by combining the phrase alignments established by the algorithm and the SL-TL phrase alignment information stored in the pair of parallel sentences.

The dynamic programming algorithm compares structures from the same language. The most similar SL structure from the bilingual corpus, that will determine the TL translation structure, is thus selected purely on SL properties. The similarity of two sentences is calculated as a weighted internal product between the two sentences, traversing both sentences in parallel from their start towards their end. The implemented method utilises the Smith-Waterman variant (Smith and Waterman, 1981).

The last step of this phase is the translation of words using the bilingual lexicon.<sup>2</sup> All translation alternatives are disambiguated during the subsequent translation phase.

## 5 Translation Phase 2: Translation equivalent selection

Issues resolved in the second phase are phrase-internal and include (i) word order within each phrase, (ii) introduction or deletion of functional words and (iii) selection of the best candidate in the case of translation ambiguities. These are resolved using the phrase-based indexing of the TL monolingual corpus.

For each phrase of the sentence being translated, the algorithm searches the TL phrase model for similar phrases. If the search is successful, all retrieved TL phrases are compared to the phrase to be translated. The comparison is based on the words included, their tags and lemmas and the morphological features.

---

<sup>2</sup> If an SL word is not included in the lexicon, it is retained in the translation in its original SL form.

1. Retrieve the relevant phrases from the TL corpus based on the head word
2. Compare the phrase with all the TL relevant phrases and store the one that scores the highest similarity score
3. For any words that the TL model cannot disambiguate, use the lemma frequency model for selecting the best translation
4. Return the new translated Phrase instance.

Figure 1. Pseudocode for Translation equivalent selection

For the purposes of the proposed methodology, the stable-marriage algorithm (Gale & Shapley, 1962) is applied for calculating the similarity and aligning the words of a phrase pair. In comparison to other relevant algorithms, the Gale-Shapley algorithm, results in potentially non-optimal solutions, but possesses the advantage of a substantially lower complexity and thus a reduced processing time.

Using the most similar TL phrase and the word alignments generated by the stable-marriage algorithm, word reordering, translation disambiguation and addition or removal of functional words is performed for each phrase of the input sentence. The final translation is produced by combining all of its translated phrases.

## 6 Developing new Language Pairs

The porting of the proposed methodology to new language pairs is straightforward. The summary presented herewith is based on the creation of a new Greek-to-Italian language pair, and is typical of porting to new TLs. Initially, the NLP tools need to be selected for the new language (tagger & lemmatiser, shallow parser). In addition, a TL monolingual corpus and a bilingual lexicon need to be provided. The following steps are then taken:

- A. Create a java wrapper class for the Italian annotation tools, and provide rules for identifying heads of phrases.
- B. Tag/lemmatise and chunk the TL corpus, which takes less than a day.
- C. Process the chunked Italian corpus to generate the phrase model. This operation is fully automated and performed off-line (e.g. for a corpus of 100 million words, approx. 1.5 days are needed).

- D. For the parallel corpus, train the PAM/PMG suite for the relevant language pair (less than 2 hours needed).

## 7 Objective Evaluation Experiments

The evaluation results reported in this article focus on the Greek – English language pair. Two datasets have been used (a development set and a test set), each of which comprises 200 sentences, with a length of between 7 and 40 words. For every sentence, exactly one reference translation has been created, by SL-language native speakers and then the translation correctness was cross-checked by TL-language native speakers.

Number of sentences	200	Source	web	
Reference translations	1	Language pair	EL-EN	
MT system	Metrics			
	BLEU	NIST	Meteor	TER
<b>PRESEMT</b>	0.3254	6.9793	0.3880	51.5330
<b>METIS-2</b>	0.1222	3.1655	0.2698	82.878
<b>Systran</b>	0.2930	6.4664	0.3830	49.721
<b>Bing</b>	0.4600	7.9409	0.4281	37.631
<b>Google</b>	0.5544	8.8051	0.4665	29.791
<b>WorldLingo</b>	0.2659	5.9978	0.3666	50.627

Table 1. Objective metrics results for PRESEMT & other MT systems (development set)

To objectively evaluate the translation accuracy, four automatic evaluation metrics have been chosen, namely BLEU (Papineni et al., 2002), NIST (NIST 2002), Meteor (Denkowski and Lavie, 2011) and TER (Snover et al., 2006). When developing the MT methodology, extensive evaluation was carried out at regular intervals (Sofianopoulos et al., 2012). The evolution of translation accuracy is depicted within Figure 2. The falling trend for TER, signifies a continuously improving translation performance. The current results for a number of MT systems for the development set are reported in Table 1. These results show that at the current stage of development the proposed approach has a quality exceeding that of WorldLingo and Systran, but is still inferior to Google and Bing. The results are particularly promising, taking into account that the proposed methodology has been developed for a substantially shorter period than the other systems, and has no language-specific information injected into it. According to an er-

ror analysis carried out, most of the errors are due to the lack of syntactic information (e.g. the inability to distinguish between object/subject). Also a point which can be improved concerns the mapping of sentence structures from SL to TL. To address this, additional experiments are currently under way involving larger monolingual corpora.

Even without this type of knowledge, the proposed methodology has shown substantial scope for improvement, as evidenced by the evolution of the objective translation metrics (cf. Figure 2). It is expected that this trend will be continued in future versions of the MT system.

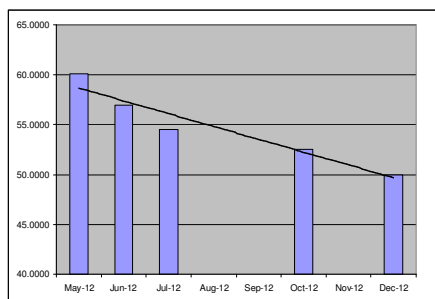


Figure 2. Evolution of translation accuracy reflected by TER scores for the PRESEMT system together with the associated trend line

<b>Number of sentences</b>	200	<b>Source</b>	web	
<b>Reference translations</b>	1	<b>Language pair</b>	EL-EN	
<b>PMG type</b>	<b>Metrics</b>			
	<b>BLEU</b>	<b>NIST</b>	<b>Meteor</b>	<b>TER</b>
<b>CRF-based</b>	0.3167	6.9127	0.3817	52.509
<b>PMG-simple</b>	0.3254	6.9793	0.3880	51.533

Table 2. Effect on PRESEMT translation accuracy of using the two distinct PMG variants

Recent activity towards improving translation accuracy has focussed on the effect of using different PMG approaches, as summarised in section 3. According to Table 2, an improvement in all four metrics is achieved using PMG-simple instead of CRF. For the limited training set defined by the parallel corpus, PMG-simple extracts more effectively the phrasing model. An improvement of approx. 3% in the BLEU score is achieved over the CRF-based system. The reduction in TER is almost 2% indicating a sizable improvement in translation quality, while

NIST and METEOR scores are improved by 1% and 1.9% respectively.

## 8 Subjective Evaluation Results

To fully evaluate translation quality, both objective and subjective evaluation have been implemented. The latter type is carried out by humans who assess translation quality.

Human evaluation is considered to be more representative of the actual MT quality (Callison-Burch, et al., 2008 & 2011), though on the other hand it is time-consuming and laborious. Furthermore, it lacks objectivity (single evaluators may not be consistent in assessing a given translation through time while two evaluators may yield completely different judgements on the same text) and must be repeated for every new test result.

For the human evaluation, for each language pair, a total of 15 language professionals were recruited, who were either language professionals, closely associated with MT tasks, or post-graduate university students in the area of linguistics. Two types of subjective evaluation were carried out. The first one involves the experts grading translations generated by the PRESEMT system regarding their adequacy and fluency. Adequacy refers to the amount of information from the SL text that is retained in the translation, based on a 1-5 scale of scores (with a score of 1 corresponding to the worst translation). Fluency measures whether the translation is well-formed, also on a 1-5 scale, with emphasis being placed on grammaticality.

The second type of subjective evaluation involves direct comparison between the translations generated by PRESEMT and by other established MT systems over the same dataset. In this case, each evaluator ranks the translations of the different systems, these systems being presented in randomised order to ensure the dependability of the feedback received.

Subjective evaluation activities were carried out during two distinct periods (namely October and December 2012), separated by two months. The purpose of implementing two sessions has been to judge the improvement in the system within the intervening period. Thus, two distinct versions of the EL-EN MT system corresponding to these two time points were used. For ref-

erence, the objective evaluation results obtained for the test sentences are listed in Table 3. In both cases, the CRF-based PMG was used since it was more mature at the time of evaluation.

A specifically-designed platform has been developed to support subjective evaluation activities<sup>3</sup>. This platform has been used to (a) collect the human evaluators' feedback for the different language pairs and (b) support the subsequent assessment of the results via statistical methods.

Number of sentences	200	Source	web	
Reference translations	1	Language pair	EL-EN	
MT system	Metrics			
	BLEU	NIST	Meteor	TER
<b>PRESEMT (phase 1)</b>	0.2627	6.2001	0.3329	60.0420
<b>PRESEMT (phase 2)</b>	0.2666	6.2061	0.3335	59.3360
<b>Bing</b>	0.4793	8.1357	0.4486	35.7220
<b>Google</b>	0.5116	8.4549	0.4580	32.6860
<b>WorldLingo</b>	0.3019	6.3799	0.3814	46.7350

Table 3. Objective metrics results for PRESEMT & other MT systems (test set)

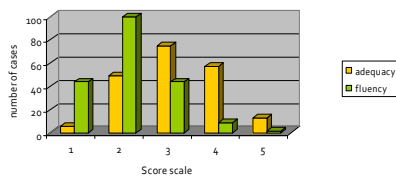


Figure 3. Histogram of adequacy and fluency over all sentences (1<sup>st</sup> human evaluation phase)

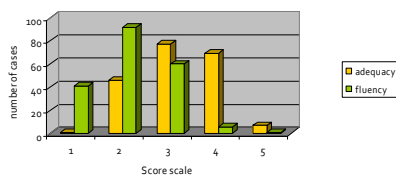


Figure 4. Histogram of adequacy and fluency over all sentences (2<sup>nd</sup> human evaluation phase)

For the proposed methodology, in phase 1 relatively low values of both adequacy and fluency

<sup>3</sup> [www.presemt.eu/presemt\\_eval/](http://www.presemt.eu/presemt_eval/)

measurements were recorded. By comparing the scores in the first and second evaluation phases (Figures 3 and 4, respectively), it can be seen that both adequacy and fluency histograms move towards higher values (notably fluency ratings with a score of 3 and adequacy ratings with scores of 3 and 4 have substantially higher frequencies). This reflects improved translation quality in the later version of the proposed MT system in comparison to the earlier one.

Number of sentences	200	Source	web	
Reference translations	1	Language pair	EL-EN	
MT system	Adequacy		Fluency	
	average	stdev.	average	stdev.
<b>PRESEMT (phase 1)</b>	3.08	0.27	2.17	0.27
<b>PRESEMT (phase 2)</b>	3.14	0.24	2.16	0.25
<b>Google</b>	4.17	0.39	3.51	0.50
<b>Bing</b>	3.75	0.77	3.02	0.61
<b>WorldLingo</b>	3.77	0.45	3.11	0.51

Table 4. Summary of measurements (in terms of average and standard deviation) for fluency and adequacy for various MT systems (test set)

In addition, in phase 2 of subjective evaluation, adequacy and fluency measurements were collected for the three operational systems used as reference systems (namely Google Translate, Bing and WorldLingo). These operational systems have higher adequacy and fluency values than PRESEMT, as indicated in Table 4. Furthermore, paired t-tests have confirmed that at a 0.99 level of significance, these three systems have statistically superior subjective measurements to the proposed methodology. To provide a reference, for the same set of 200 sentences, objective metrics are shown in Table 3 for each system. As can be seen the relative order of the systems in the subjective evaluations (in terms of adequacy and fluency) is confirmed by the objective measurements.

A second subjective evaluation focused on ranking comparatively the translations of the four studied MT systems. Evaluators were presented with the outputs of the four systems in randomized order, to conceal the identity of each system. The evaluators were requested to order the four translations from higher to lower quality (with 1 denoting the more accurate translation.

To transform this ranking into a single score, the individual rankings per evaluator have been accumulated and normalized over the number of evaluators. Then the representative scoring has been defined as a weighted sum of frequency of a system being ranked as first, second, third and fourth best over all evaluators, by multiplying with weights of 40, 30, 20 and 10 respectively. The average scores of the proposed methodology were the lowest, followed by the ranking results for WorldLingo. The results of Bing and Google are comparable with the Google results giving the best results. A statistical analysis was carried out using paired t-tests for all six pairings of the four systems being studied. This has confirmed that the differences in subjective scores are statistically significant at a level of 0.95.

To summarise, subjective evaluation has shown that the PRESEMT methodology has an inferior translation performance in terms of subjective measurements to the three operational systems. This can be justified as the proposed methodology refrains from utilising language-specific information as a priori grammatical knowledge. Inferior translations also reflect the much shorter development time available as well as the very limited amount of expensive resources provided. The effect on translation quality of using pre-existing tools (to ease portability to new language pairs) needs to be stressed, as no modification of these tools was performed to remedy systematic shortcomings identified. For the newer MT versions now available, a new round of subjective evaluations is planned. It has been observed that improvements in objective metrics are followed by improved subjective evaluation performance. Thus, for these new versions, an improved accuracy is expected.

## 9 Discussion

In the present article the principles and implementation of a novel language-independent MT methodology have been presented. This methodology draws on information from a large TL monolingual corpus and a very small bilingual one. The overwhelming majority of linguistic information is extracted in an automated manner using pattern recognition techniques.

Two types of evaluation have been reported, these concerning objective and subjective

evaluations. Experimental results using objective metrics through a period of time have indicated a rising trend in terms of translation quality. Also, it has been shown that by introducing a new phrasing model for the sentences to be translated a substantial improvement is achieved. Subjective evaluation activities have indicated a higher translation accuracy achieved by other MT systems. A limiting factor for the PRESEMT methodology is admittedly the requirement for portability to new language pairs. This leads to the extraction of knowledge from texts via algorithmic means and the adoption of already existing linguistic tools, without modifications.

On the other hand, subsequent versions of the proposed MT system have shown a trend of improving translation accuracy. In this respect, objective evaluation results are promising, especially taking into account the fact that for several aspects, scope for improvement has been identified. This includes the revision of the structure selection phase, where smaller sub-sentential structures need to be combined to improve generalisation. In addition, improvements in the bilingual corpus compilation procedure need to be studied. The results of these ongoing experiments will be reported in the future.

## References

- Paul E. Black. 2005. *Dictionary of Algorithms and Data Structures*. U.S. National Institute of Standards and Technology (NIST).
- Chris Callison-Burch, Cameron Fordyce, Philipp Koehn, Christof Monz and Josh Schroeder. 2009. *Further Meta-Evaluation of Machine Translation*. Proceedings of the WMT-08 Workshop, Columbus, Ohio.
- Chris Callison-Burch, Philip Koehn, Christof Monz, Omar F. Zaidan. 2011. *Findings of the 2011 Workshop on Statistical Machine Translation*. Proceedings of the 6<sup>th</sup> Workshop on Statistical Machine Translation, Edinburgh, UK, pp. 22–64.
- Jaime Carbonell, Steve Klein, David Miller, Michael Steinbaum, Tomer Grassiany and Jochen Frey. 2006. *Context-Based Machine Translation*. Proceedings of the 7<sup>th</sup> AMTA Conference, Cambridge, MA, USA, pp. 19-28.
- Michael Carl, Maite Melero, Toni Badia, Vincent Vandeghinste, Peter Dirix, Ineke Schuurman, Stella Markantonatou, Sokratis Sofianopoulos,

- Marina Vassiliou and Olga Yannoutsou. 2008. METIS-II: Low Resources Machine Translation: Background, Implementation, Results and Potentials. *Machine Translation*, 22 (1-2):pp. 67-99.
- Helena M. Caseli, Maria das Graças V. Nunes and Mikel L. Forcada. 2008. Automatic Induction of Bilingual resources from aligned parallel corpora: Application to shallow-transfer machine translation. *Machine Translation*, 20:pp. 227-245.
- Michael Denkowski and Alon Lavie. 2011. Meteor 1.3: *Automatic Metric for Reliable Optimization and Evaluation of Machine Translation Systems*. EMNLP 2011 Workshop on Statistical Machine Translation, Edinburgh, UK, pp. 85-91.
- Ioannis Dologlou, Stella Markantonatou, George Tambouratzis, Olga Yannoutsou, Athanasia Fourla and Nikos Ioannou. 2003. *Using Monolingual Corpora for Statistical Machine Translation: The METIS System*. Proceedings of the EAMT- CLAW 2003 Workshop, Dublin, Ireland, pp. 61-68.
- Richard O. Duda, Peter E. Hart and David G. Scott. 2001. *Pattern Classification (2<sup>nd</sup> edition)*. Wiley Interscience, New York, U.S.A.
- David Gale and Lloyd S. Shapley. 1962. College Admissions and the Stability of Marriage. *American Mathematical Monthly*, 69:pp. 9-14.
- John Hutchins. 2005. Example-Based Machine Translation: a Review and Commentary. *Machine Translation*, 19:pp. 197-211.
- Alexandre Klementiev, Ann Irvine, Chris Callison-Burch and David Yarowsky. 2012. *Toward Statistical Machine Translation without Parallel Corpora*. Proceedings of EACL2012, Avignon, France, 23-25 April, pp. 130-140.
- Philip Koehn. 2010. *Statistical Machine Translation*. Cambridge University Press, Cambridge.
- John Lafferty, Andrew McCallum and Fernando Pereira. 2001. *Conditional Random Fields: Probabilistic Models for Segmenting and Labelling Sequence Data*. Proceedings of ICML 2011, Bellevue, Washington, USA, pp. 282-289.
- Harry Mairson. 1992. The Stable Marriage Problem. *The Brandeis Review*, 12:1.
- Stella Markantonatou, Sokratis Sofianopoulos, Olga Giannoutsou and Marina Vassiliou. 2009. Hybrid Machine Translation for Low- and Middle- Density Languages. *Language Engineering for Lesser-Studied Languages*, S. Nirenburg (ed.), IOS Press, pp. 243-274.
- NIST 2002. *Automatic Evaluation of Machine Translation Quality Using n-gram Co-occurrences Statistics*.
- Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. 2002. *BLEU: A Method for Automatic Evaluation of Machine Translation*. Proceedings of the 40<sup>th</sup> ACL Meeting, Philadelphia, USA, pp. 311-318.
- Jan Pomikálek and Pavel Rychlý. 2008. *Detecting co-derivative documents in large text collections*. Proceedings of LREC2008, Marrakech, Morocco, pp.1884-1887.
- Prokopis Prokopidis, Byron Georgantopoulos and Harris Papageorgiou. 2011. *A suite of NLP tools for Greek*. Proceedings of the 10<sup>th</sup> ICGL Conference, Komotini, Greece, pp. 373-383.
- Felipe Sanchez-Martinez and Mikel L. Forcada. 2009. Inferring Shallow-transfer Machine translation Rules from Small Parallel Corpora. *Journal of Artificial Intelligence Research*, 34:pp. 605-635.
- Helmut Schmid. 1994. *Probabilistic Part-of-Speech Tagging Using Decision Trees*. Proceedings of International Conference on New Methods in Language Processing, Manchester, UK, pp. 44-49.
- Temple F. Smith and Michael S. Waterman. 1981. Identification of Common Molecular Subsequences. *Journal of Molecular Biology*, 147:195-197.
- Matthew Snover, Bonnie Dorr, Richard Schwartz, Linnea Micciulla and John Makhoul. 2006. *A Study of Translation Edit Rate with Targeted Human Annotation*. Proceedings of the 7<sup>th</sup> AMTA Conference, Cambridge, MA, USA, pp. 223-231.
- Sokratis Sofianopoulos, Marina Vassiliou and George Tambouratzis. 2012. *Implementing a language-independent MT methodology*. Proceedings of the 1<sup>st</sup> Workshop on Multilingual Modeling (held within the ACL-2012 Conference), Jeju, Republic of Korea, pp.1-10.
- Jinsong Su, Hua Wu, Haifeng Wang, Yidong Chen, Xiaodong Shi, Huailin Dong and Qun Liu. 2011. *Translation Model Adaptation for Statistical Machine Translation with Monolingual Topic Information*. Proceedings of the 50<sup>th</sup> ACL Meeting, Jeju, Republic of Korea, pp. 459-468.
- George Tambouratzis, Fotini Simistira, Sokratis Sofianopoulos, Nikos Tsimboukakis and Marina Vassiliou. 2011. *A resource-light phrase scheme for language-portable MT*. Proceedings of the 15<sup>th</sup> EAMT Conference, Leuven, Belgium, pp. 185-192.